STIC Search Report
Biotech-Chem Library

TO: Jennifer Graser
Location: 3b09 / 3c18
Wednesday, April 07, 2004
Art Unit: 1645
Phone: 272-0858
Serial Number: 09 / 844281

From: Jan Delaval
Location: Biotech-Chem Library
Rem 1A51
Phone: 272-2504

jan.delaval@uspto.gov

Search Notes

118973

From: Hutzell, Paula
Sent: Wednesday, April 07, 2004 3:01 PM
To: Graser, Jennifer; STIC-Biotech/ChemLib
Subject: RE: rush search

approved

-----Original Message-----
From: Graser, Jennifer
Sent: Wednesday, April 07, 2004 12:26 PM
To: Hutzell, Paula
Subject: rush search
Importance: High

Hi Paula,

Could you please authorize the following rush search for an amendment which is due?

Thanks,
Jennifer

STIC:
Please search **SEQ ID NO: 1** from 09/844,281 in pending and commercial databases.

Thanks,
Jennifer Graser
REMSEN
E03 B09/ 3C18 mailbox
AU 1645
272-0858

RECEIVED
APR - 7 2004
(STIC)

Searcher: _____
Phone: _____ 2250Y
Location: _____
Date Picked Up: ___4 7___
Date Completed: ___4 7___
Searcher Prep/Review: _____
Clerical: _____ c o
Online time: _____ 10

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: ___✓_____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: ___✓_____
WWW/Internet: _____
Other (specify): _____

OM protein - protein search, using sw model

```
                GenCore   version 5.1.6
          Copyright  (c) 1993 - 2004  Compugen Ltd.
```

Run on: April 7, 2004, 17:29:55 ; Search time 18 Seconds
(without alignments)
2409.691 Million cell updates/sec

Title: US-09-844-281-1
Perfect score: 4202
Sequence: 1 AGKSFPDVPAGHWAEGSINY.........ITSEIGSQAVHVNVLNNPNL 833

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	
1	4202	100.0	862	1	SLA2_BACAA	bacillus an	
2	2833.5	67.4	874	1	SLAP_BACLI	bacillus li	
3	719	17.1	814	1	SLA1_BACAA	bacillus an	
4	529	12.6	531	1	YQ42_BACAN	bacillus an	
5	348	8.3	1176	1	SLAP_BACSH	bacillus sp	
6	265.5	6.3	762	1	SLAP_ACEKI	acetogenium	
7	252.5	6.0	941	1	GUN_BACS6	bacillus sp	
8	230	5.5	1325	1	YDEK_ECOLI	escherichia	
9	223.5	5.3	2003	1	YDBA_ECOLI	escherichia	
10	219	5.2	1116	1	SLPH_BRECH	brevibacill	
11	217	5.2	1655	1	OMPB_RICCN	rickettsia	
12	211	5.0	1053	1	SLPM_BACBR	bacillus br	
13	211	5.0	1645	1	OMPB_RICTY	rickettsia	
14	207.5	4.9	2358	1	YEEJ_ECOLI	escherichia	
15	207.5	4.9	2660	1	YEEJ_ECO57	escherichia	
16	206.5	4.9	1654	1	OMPB_RICRI	rickettsia	
17	206	4.9	1698	1	41_DROME	drosophila	
18	200	4.7	1643	1	OMPB_RICPR	rickettsia	
19	199.5	4.7	939	1	SLAP_CAMPE	campylobact	
20	197.5	4.7	1608	1	HLYA_SERMA	serratia ma	
21	195.5	4.6	1848	1	CBPA_CLOCL	clostridium	
22	194.5	4.6	1656	1	OMPB_RICJA	rickettsia	
23	194.5	4.6	1902	1	P2P_LACPA	lactobacill	
24	194	4.6	1300	1	120K_RICRI	rickettsia	
25	193	4.6	917	1	SLAP_THETH	thermus the	
26	189.5	4.5	1183	1	CNA_STAAU	staphylococ	
27	189	4.5	1953	1	BIGA_SALTY	salmonella	
28	188.5	4.5	1276	1	PMP6_CHLPN	chlamydia p	
29	187.5	4.5	1861	1	APU_THETU	t amylopull	
30	185.5	4.4	1577	1	HLYA_PROMI	proteus mir	
31	183.5	4.4	1087	1	XYNX_CLOTM	clostridium	
32	183	4.4	2249	1	OMPA_RICRI	rickettsia	
33	183	4.4	4705	1	FAT2_DROME	drosophila	
34	181.5	4.3	1228	1	SLAP_BACST	bacillus st	P35825 bacillus st
35	181.5	4.3	1902	1	P2P_LACLC	lactococcus	P15293 lactococcus
36	180	4.3	1286	1	AIDA_ECOLI	escherichia	Q03155 escherichia
37	180	4.3	1569	1	YPJA_ECOLI	escherichia	P52143 escherichia
38	179.5	4.3	1726	1	MSP1_PLAFC	plasmodium	P04934 plasmodium
39	179.5	4.3	1726	1	MSP1_PLAFP	plasmodium	P50495 plasmodium
40	179	4.3	1664	1	SLP1_CLOTM	clostridium	Q06852 clostridium
41	178.5	4.2	2021	1	OMPA_RICCN	rickettsia	Q52657 rickettsia
42	176.5	4.2	447	1	ANCA_CLOTM	clostridium	Q06848 clostridium
43	176.5	4.2	1902	1	P1P_LACLC	lactococcus	P16271 lactococcus
44	174	4.1	1025	1	SLAP_CAUCR	caulobacter	P35828 caulobacter
45	172	4.1	1020	1	BCA_STRAG	streptococc	Q02192 streptococc

ALIGNMENTS

RESULT 1
SLA2_BACAA
ID SLA2_BACAA   STANDARD;   PRT;   862 AA.
AC P94217;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE S-layer protein EA1 precursor.
GN EAG OR BA0887.
OS Bacillus anthracis (strain Ames), and
OS Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=198094, 1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames;
RX MEDLINE=22608414; PubMed=12721629;
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria.";
RL Nature 423:81-86(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sterne / 9131;
RX MEDLINE=97260111; PubMed=9106206;
RA Mesnage S., Tosi-Couture E., Mock M., Gounon P., Fouet A.;
RT "Molecular characterization of the Bacillus anthracis main S-layer
RT component: evidence that it is the major cell-associated antigen.";
RL Mol. Microbiol. 23:1147-1155(1997).
-!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
    of proteins which coat the surface of bacteria.
-!- SUBCELLULAR LOCATION: Cell wall.
-!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.

DR EMBL; AE017027; AAP24884.1; -.
DR EMBL; X99724; CAA68063.1; -.
DR TIGR; BA0887; -.
DR InterPro; IPR001119; SLH.

RESULT 1 — S-LAYER PROTEIN EA1

```
DR   Pfam; PF00395; SLH; 3.
KW   Signal; Repeat; Cell wall; S-layer; Complete proteome.
FT   SIGNAL        1     29       POTENTIAL.
FT   CHAIN        30    862       S-LAYER PROTEIN EA1.
FT   DOMAIN       34     76       SLH 1.
FT   DOMAIN       95    136       SLH 2.
FT   DOMAIN      157    197       SLH 3.
SQ   SEQUENCE    862 AA;  91362 MW;  CB16B202F62CCCA0 CRC64;

Query Match             100.0%;   Score 4202;   DB 1;   Length 862;
Best Local Similarity   100.0%;   Pred. No. 1.3e-175;
Matches  833;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1  AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE   60

Qy     61  NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV  120

Qy    121  NGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYGK  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  NGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYGK  180

Qy    181  KDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVTLGGK  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  KDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVTLGGK  240

Qy    241  VAPNKDLTVKVKNQSFVTKFVYEVKKLAVEKLTPDDDRAGQAIAFKLNDEKGNADVEYLN  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  VAPNKDLTVKVKNQSFVTKFVYEVKKLAVEKLTPDDDRAGQAIAFKLNDEKGNADVEYLN  300

Qy    301  LANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLTVSNTG  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  LANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLTVSNTG  360

Qy    361  IITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLNKLVA  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  IITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLNKLVA  420

Qy    421  TIAGEDKVVDPGSISIKSSNHGIISVNNYITAEAAGEATLIIKVGDVTKDVKFKVTTDS   480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  TIAGEDKVVDPGSISIKSSNHGIISVNNYITAEAAGEATLIIKVGDVTKDVKFKVTTDS   480

Qy    481  RKLVSVKANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVV  540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  RKLVSVKANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVV  540

Qy    541  TTDSGSIGTKTIGVTGNDVGEGTVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQY  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  TTDSGSIGTKTIGVTGNDVGEGTVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQY  600

Qy    601  GQSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVT  660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  GQSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVT  660

Qy    661  GKTPGKVDIHLTKNGATAGKATVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLELEK  720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  GKTPGKVDIHLTKNGATAGKATVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLELEK  720

Qy    721  SNLDDIVKGINLTKETQHKVRVVKSGAEQGKLYLDRNGDAVFNAGDVKLGDVTVSQTSDS  780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  SNLDDIVKGINLTKETQHKVRVVKSGAEQGKLYLDRNGDAVFNAGDVKLGDVTVSQTSDS  780

Qy    781  ALPNFKADLYDTLTKYTDKGTLVFKVLKDKDVITSEIGSQAVHVNVLNNPNL          833
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  ALPNFKADLYDTLTKYTDKGTLVFKVLKDKDVITSEIGSQAVHVNVLNNPNL          833
```

RESULT 2 — SLAP_BACLI

```
KW   Signal; Cell wall; S-layer; Repeat.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31    874       S-LAYER PROTEIN.
FT   DOMAIN       33     93       SLH 1.
FT   DOMAIN       94    155       SLH 2.
FT   DOMAIN      156    217       SLH 3.
SQ   SEQUENCE    874 AA;  92734 MW;  EFADCA4FF27D32AF CRC64;

Query Match             67.4%;   Score 2833.5;   Pred. No. 3.6e-116;   DB 1;   Length 874;
Best Local Similarity   67.3%;
Matches  569;  Conservative   98;  Mismatches  156;  Indels   23;  Gaps    7;
```

```
Qy      1  AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE   60
           |||||||||||||||:|||||||||||:||||||||||||||||||||||||||||||||
Db     30  AGKSFPDVPAGHWAEDSINYLVDKGAIVGKPDGTYGPTESIDRASAAVIFTKILNLPVDE   89

Qy     61  NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV  120
           ||||||||||||||||||||||||||||||||:||||||||||||||||||:||||:|||
Db     90  NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKDNFYPEGKIDRASFASMLVGAYNLKEKV  149

Qy    121  NGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYGK  180
           :|:|||.|:||:.|||||||||:||||||||||||||||||||||||||||||||||:||:
Db    150  DGTLVTKFDDLLRGHWGEEKANILVNLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYAK  209

Qy    181  KDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVTLGGK  240
           .|:.:.:||.::::||:.|||||||:||::|:||||||:||:|:|||.||.|||||||.||
Db    210  PENSDAKVTNVAATEPTQLTLTGTGLNKLTAEDVTLEGNKAIALEASKDGKSAVVTLSGK  269

Qy    241  VAPNKDLTVKVKNQSFVTKFVYEVKKLAVEKLTPDDDRAGQAIAFKLNDEKGNADVEYLN  300
           :|||:.|.|||.:.:|:.|:|YEVKKL.VE:LT.DDDRA.|A::FKLND:KGNAD:EYL:
Db    270  IAPNKELPVKVKGNTFIVKYVYEVKKLRVEQLTFDDDRADQAVVFKLNDKKGNADIEYLD  329

Qy    301  LANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLTVSNTG  360
           :|.HDVKFVANNLDG:PANIFEGG.A.STTGKLAVGI.:G.YKVEVQVTKRGG..TVSNTG
Db    330  IAGHDVKFVANNLDGTPANIFEGGTARSTTGKLAVGIAEGKYKVEVQVTKRGGIITVSNTG  389

Qy    361  IITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLNKLVA  420
           II.VKNLD..A:AIK:VVFA:D:D..G-VNY..KLSG.DF::NS:.:VVGEKA.::K:VA
Db    390  IIEVKNLDAEATAIKDVVFAVDTDKAG-VNYAKPLSGTDFTINSKTIVAGEKAGIHKVVA  448
```

RESULT 2
SLAP_BACLI
ID   SLAP_BACLI   STANDARD;   PRT;   874 AA.

```
Qy  421 TIAGEDKVVDPGSISIKSSNHGIISVVNNYITAERAAGEATLTIKVGDVTIKDVKFKVTTDS 480
            |:|||  |:|||   |:||:|  |:|:||    ||:|||| :|||||| ||:|:||:||
Db  449 QINKENKVVDPSAISLKSSNPGVISVKNGEIKAEAAGSATLTVKVGDVTKTFDFVVKTDT 508

Qy  481 RKLVSVKANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVV 540
            ||| ||||||:|||| |:|:|||||||||||||||||||:|||||| |:|||||  |||:
Db  509 RKLJTTVKANPDQLKVVDGKELPVTFVTTDQYGDPFGANSGAIKEVFPQTGVVKV--LDVT 566

Qy  541 TTDSGSIGTKTIGVTGNDVGETVHFQNGN--GATLGSLYNVTEGNVAPK--NPELVSK 595
            |||:||||| :|:||:|||| ||||||| :| |:|||  ||:|:|:    |||||||
Db  567 TTNEGSIGTSSIKVKGENVGAGTIHFQNPNASGEGYGSLHVEVTKSNIGHEAPRLELVSK 626

Qy  597 VGQYQGSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEGYEYVESKNLAVADAKIVGNK 656
            :|| :|  ||||| ||:|  ||||:|| ||| |:|   ||||| ||:|||||||||:||:|
Db  627 AGQKGEAADT--TLGAGNTVAVQLSNVTTEGVYADADADLAGYEFRVGNDKIASAKIEGKT 684

Qy  657 VVVTGKTPGKVDIHLHKNGATAGKATVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVL 716
            ::|||||.||:|: |:|||||||:||:|:|:|||  ||||||:|| ||||:||:|||::||
Db  685 LKVTGKTAGVTDVILFKDGATAGHATITVTQENIQTSVKFKDVEVEQFENRKVNIDRVL 744

Qy  717 ELEKSNLDDIVKGINLTKETCQHKVRVVKSGAEQGKLYLDRNGDAVPNAGDVKLGDVTVSQ 776
            :|:||  |||::|| ||| |:|||||:|:||||:||||||||:|||  ||||:||||.||:
Db  745 DVVKSDKDDVLNGIKLNISTEHKVRIVDEGTEQGKVYLDRNDNATFDGNDVALGVTAVK 804

Qy  777 TSDSA-------LPNF-------KADLYDTLTTKYTDKGTIVFKVLKDKDVITSEIGSQA 822
            :||:|       |.|:       |||||||| |:|:||:|||||: ||| ||::|:||:|
Db  805 SNDTVSKEGNDLFKFLTDETATNKNDVFKGVTTAFGDKGTVIPFKVMKDRVAPTTEYGTKA 864

Qy  823 VHVNVL 828
            |::|:|
Db  865 VTINVI 870
```

```
RL   J. Bacteriol. 177:614-620(1995).
CC   -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
CC       of proteins which coat the surface of bacteria.
CC   -!- SUBCELLULAR LOCATION: Cell wall.
CC   -!- PTM: PROBABLY GLYCOSYLATED.
CC   -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
```

```
DR   EMBL; AE017027; AAP24883.1; -.
DR   EMBL; 236946; CAA85408.1; -.
DR   PIR; I40048; I40048.
DR   TIGR; BA0885; -.
DR   InterPro; IPR003343; Big_2.
DR   InterPro; IPR008964; Invasin_intimin.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF02368; Big_2; 1.
DR   Pfam; PF00395; SLH; 3.
KW   Signal; Glycoprotein; Repeat; Cell wall; S-layer; Complete proteome.
FT   SIGNAL        1     29
FT   CHAIN        30    814       S-LAYER PROTEIN SAP.
FT   DOMAIN       33     93       SLH 1.
FT   DOMAIN       94    154       SLH 2.
FT   DOMAIN      155    213       SLH 3.
SQ   SEQUENCE   814 AA;  86620 MW;  C1638D26A1C6B101 CRC64;
```

```
Query Match             17.1%;  Score 719;  DB 1;  Length 814;
Best Local Similarity   29.5%;  Pred. No. 2.3e-24;
Matches  229;  Conservative  126;  Mismatches  302;  Indels  118;  Gaps  27;

Qy    1 AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGYGPTESIDRASAAVIFTKILNLPVDE 60
          |||:||||||.||. .:||||||:|||.|| ||||:|:||||||:||: |:||||||:|
Db   30 AGKTFPDVPADHWGIDSINYLVEKGAVKGNDKGMFEPGKELTRAEAATMMAQILNLPIDK 89

Qy   61 NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV 120
          :|:|||.|: ::|::::||||||||||:|||| |:|:|.||||.|:||:||:||:||:||
Db   90 DAKPSFADSGQWYTPFIAAVEKAGVIKGTG-NGFEPNGKIDRVSMASLLVEAYKLDTKV 148

Qy  121 NGELVTTFEDLLD-HWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYG 179
          || .|| |:||   :||::||||::.||||||||.:|||.|::|:||||||||:|||::.
Db  149 NGTPATKFKDLETLNWGKEKANILVELGISVGTGDQWEPKKTVTKAEAAQFIAKTDKQFG 208

Qy  180 ----KKDNAQAYVT--DVKVSEPKLTLTGTGLDKLSADDVTL---EGDKAVAIEA- 226
              ||:||  :|  |.|||:|  .:  :|    |:|||||:|   |||||||:|
Db  209 TEAAKVESARKAVTTQKVEVKFSK-------AVEKLTKEDIKVTNKANNDKVLVKEVTL 259

Qy  227 STDGTSAVVTLGGKVAPNKDLTVKV----KNQSFVTKFVYEVKKLAVEKLTFDDDRAGQA 282
          |.|:||:|.|  .|  .|| |||:|    ||:|:||:||.||::|:||:|||:::.|
Db  260 SEDKKSATVELYSNLAAKQTYTVDVNKVGKTEVAVGSLEAKTIEMADQTVVADEP--TA 316

Qy  283 IAFKLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDY 342
          :.|:|:|| |.:||.::    :|.:|.   ..|||||||||||||||||:|||||:||:|
Db  317 LQFTVKDENGTEVV-----SPEGIEFV------AEKINAKGEITLAKGTS 358

Qy  343 KVEVQVTKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFAL 402
          ..:..||:.||:|||:||||||||:||||||||||||||||||||||||||||||||||:
Db  359 TTVKAVYKDGDGKVVABSKEVKVSABGAAVASISNWTVAEQNKAD-----FTSKDFKQ 410

Qy  403 NSQNLVVGEKASLNKLVATIAGEDKVVDPGSISIKSSNH--GIISVVNNYITAEAAGEAT 460
          |:::::.|.: :|::.|.|. :|:.:||:.| ||||||:  ||:|||||||||||||||||
Db  411 NNK-VYEGDNAYVQ---VELKDQFNAVTTGKVEYESLNTEVAVVDKATGKVTVLSAGKAP 466

Qy  461 LTIKVGD------VTKDVKFKV-TTDSRKLVSVKANPDKLQVVQNKTLPVTFVTTDQYGD 513
          :.:.|||      |.||||||| ||||.|||||||||||||||||||||||||||||||||
Db  467 VKVTVKDSKGKELVSKTVEIEAFAQKAMKEIKLEKTNVALSTKDVTDLKVKAPVLDQYGK 526

Qy  514 PFGANTAAIKEVLPKTG-VVAEGGLDVVTTDSGSIGTKTIGVTGNDVGEGTVHF-QNGN 570
```

RESULT 3
SLA1_BACAA

```
ID   SLA1_BACAA     STANDARD;      PRT;   814 AA.
AC   P49051;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   S-layer protein sap precursor (Surface layer protein) (Surface array
DE   protein).
GN   SAP OR BA0885.
OS   Bacillus anthracis (strain Ames), and
OS   Bacillus anthracis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=198094, 1392;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ames;
RX   MEDLINE=22608414; PubMed=12721629;
RA   Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA   Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA   Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA   Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA   DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA   Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA   Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA   Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA   Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA   Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA   Fraser C.M.;
RT   "The genome sequence of Bacillus anthracis Ames and comparison to
RT   closely related bacteria.";
RL   Nature 423:81-86(2003).
RN   [2]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 30-40; 209-218; 281-291 AND
RP   571-580.
RC   STRAIN=Sterne;
RX   MEDLINE=95138020; PubMed=7836294;
RA   Etienne-Toumelin I., Sirard J., Duflot E., Mock M., Fouet A.;
RT   "Characterization of the Bacillus anthracis S-layer: cloning and
RT   sequencing of the structural gene.";
```

```
Db  527 EFTAPVTV-.KVLDKDGKELKEQKLEAKYVNKELV---LNAAGQEAGNYTVVLTAKSGE 580
         :|||          |. |:.||||.||:.|||:
Qy  571 GATLGSLYVNVTEGNVAFKNFELVSKVGQYGQSPDTKLDLNVSTTVEYQLSKYTSDRVYS 630
          :..|||:  |
Db  581 KEAKATLALEL-KAPGAFSKFEV---------------RGLEKELDKYVTEENQK 619

Qy  631 DPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVD-----------IHLTKNGATA 678
         :|.| |            |          |   |  |           ||| |.|||
Db  620 NAMTVSVLPVDANGLVLKGAEAABLKVTTNKEGKEVDATDAQVTVQNNSVITVGQGAKA 679

Qy  679 GKA-TVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLEEKSNLDDIVKGINL 732
         |:| ||: | |:.|.        |.      |.|| | : :  |:|:.||
Db  680 GETYKVTVVLDGKLITTHSFKVVDTAPTAK-----GLAVEFTSTSLKEVAPNADL 729
```

RESULT 4
YQ42_BACAN
ID YQ42_BACAN   STANDARD;   PRT;   531 AA.
AC Q9RMZ0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical cell-wall amidase pXO2-42 precursor (EC 3.5.1.-).
GN PXO2-42.
OS Bacillus anthracis.
OG Plasmid pXO2.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RA Okinaka R.T., Cloud K., Hampton O., Hill K.K., Keim P., Lamke G.,
RA Kumano S., Manter D., Martinez Y., Svensson R., Tatum L.R.,
RA Brown A.E., Jackson P.J.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: Cell wall (Potential).
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE N-
CC   ACETYLMURAMOYL-L-ALANINE AMIDASE FAMILY 3.
CC -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AF188935; AAF13647.1; -.
DR InterPro; IPR002508; Amidase_3.
DR InterPro; IPR001119; SLH.
DR Pfam; PF01520; Amidase_3; 1.
DR Pfam; PF00395; SLH; 3.
DR SMART; SM00646; Ami_3; 1.
DR PROSITE; PS01072; SLH_DOMAIN; FALSE_NEG.
KW Hypothetical protein; Hydrolase; Cell wall; Repeat; S-layer; Signal;
KW Plasmid.
FT SIGNAL   1   26   POTENTIAL.
FT CHAIN   27   531   HYPOTHETICAL CELL-WALL AMIDASE PXO2-42.
FT DOMAIN   33   72   SLH 1.
FT DOMAIN   91   132   SLH 2.
FT DOMAIN   153   193   SLH 3.
FT DOMAIN   400   521   AMIDASE.
SQ SEQUENCE   531 AA; 58864 MW; 9A171DC4ED05CA78 CRC64;

Query Match   12.6%; Score 529; DB 1; Length 531;
Best Local Similarity   33.2%; Pred. No. 2.4e-16;
Matches 158; Conservative 68; Mismatches 184; Indels 66; Gaps 17;

Qy   3 KSFPPVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENA 62
         | | || |: .|||.|.||| |||.||| ||||:||::|:||:|| :|| | ||
Db   31 KTFTDVP--NWAQQSVNYLMKK-AILDGKPDGTFSSSEKIDRGSAAKLMAMVLGIQINKQA 87
```

```
Qy   63 QPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASPASMLVSAYNLKDKVNG 122
         :|||||||.  ||.||| ||.||.||||:|| |: :|.  |||||:|: |:  |
Db   88 KPSFQDAKNHWASPYIAAVEKAGVYGDGSGNFNPSKDIDRASMASMLVEAYKLNNRIIG 147

Qy  123 ELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKYGKKD 182
         |:  .|:|| |.|||...|.|  .|  ||.  ||||.| ..:|| |||.|:.|| |.||
Db  148 DLPTQFEDLKGHWGAKLANALVALGISKGFTGDGWKPNGIVTRAEAVQFIAQTDMK--KAD 205

Qy  183 NAQAYVTD---VKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVT-LG 238
         .|.||.  :  |.||  |  |.:| .| |. |..|.|.  :: : ||:
Db  206 TSKRMVMNRHFITYHQP--SLSSGVTSNQHAPQI-----IVVKEQRADGWIKIVTNIG 256

Qy  239 GKVAP--NKDLTVKVNQSFVT--KPVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEKG-- 292
         |.|||  |||:|||.||||.   |.| :.:|||: |  |||.|...||.|| :||||:
Db  257 DKWTPLYEKRETI--HSTFTTYPEASHSSKVLGT-------HSPQTVT--VIEEKGSW 303

Qy  293 ----NADVEYLNLANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGD--YKVE 345
            :|| : |.|.|:.|  || |||.:.|| | . : ...|||  ||.||:||.  |||
Db  304 IRIRTNASFQWLDKNQLTLPKKQNNFLEGKTIIIDPGHGGIDGGHKGIYMNESPVVYDTA 363

Qy  346 VQVTKRGGLTVSNTGIITVKNLDTPA-----SAIKNVVFALDADNDGVVN---------- 390
         |.||:.|  |  |||: |.|  | ||       |:|| |  | ||.|.
Db  364 VRVQKLPAQKTPFTALLTRDAYSRPGKNATDSLGKRVEFAKKNKGDIFVSIHANGPNGNA 423

Qy  391 ------YGSKLSGKDFALNSQNLVVGEKASLNKLVATIAGEDKVVDPGSISIKSSN 440
               ||| | |||||.:.|::|.||||| |:|| ||:|:| || ||||:||:|::|
Db  424 HGTETFYYKAPTQKSNPYVNDSRIIAEKIQ-KRLITALQTRDRGVKIGNLIYVLREN 478
```

RESULT 5
SLAP_BACSH
ID SLAP_BACSH   STANDARD;   PRT;   1176 AA.
AC P3853?;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Surface-layer 125 kDa protein precursor.
OS Bacillus sphaericus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1421;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2362;
RX MEDLINE=89327128; PubMed=2666389;
RA Bowditch R.D., Baumann P., Yousten A.A.;
RT "Cloning and sequencing of the gene encoding a 125-kilodalton
RT surface-layer protein from Bacillus sphaericus 2362 and of a related
RT cryptic gene.";
RL J. Bacteriol. 171:4178-4188(1989).
CC -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
CC   of proteins which coat the surface of bacteria.
CC -!- SUBCELLULAR LOCATION: Cell wall.
CC -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M28361; AAA50256.1; -.
DR PIR; A33856; A33856.
DR InterPro; IPR001119; SLH.
DR Pfam; PF00395; SLH; 3.
DR PROSITE; PS01072; SLH_DOMAIN; 2.
KW Signal; Cell wall; S-layer; Repeat.
FT SIGNAL   1   30   POTENTIAL.
FT CHAIN   31   1176   SURFACE LAYER 125 kDa PROTEIN.
FT DOMAIN   32   91   SLH 1.
FT DOMAIN   92   151   SLH 2.
FT DOMAIN   152   210   SLH 3.
```

SQ SEQUENCE 1176 AA; 125225 MW; 1A9A458EF433788C CRC64;

Query Match 8.3%; Score 348; DB 1; Length 1176;
Best Local Similarity 24.2%; Pred. No. 4.7e-08;
Matches 236; Conservative 123; Mismatches 307; Indels 310; Gaps 55;

```
Qy    13  WAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQPSFKDAK-N   71
              :: :::       |||        :  : :  ||||| |      : ||||
Db    42  YAKEAVQSLVDAGVIQGDANGNFNPLKTISRAEAATIFTNALELEAEGDV--NFKDVKAD   99

Qy    72  IWSSKYIRAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGELVTTFED-  130
              :|:  :       |||::: :|| :    :   :|:  |:||:  : |:|
Db   100  AWYYDAIAATVENGIFEGVSATEFAPNKQLTRSEAAKILVDAFELEGE--GDL-SEFADA  156

Qy   131  -LLDHWGEEKANILINLGISVGT--GGK--WEPNKSVSRAEAAAQFIALTDKKYGKKDNAQ  185
              :  :  :      :     :   |:|:::
Db   157  STVKPWAKSYLEIAVANGVIKGSEANGKTNLNPNAPITRQDFAVVFSRT----------  205

Qy   186  AYYTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEAST------DGTSAVVTLGG  239
              ::|      |||       :|:|||:|:          :
Db   206  ---------------IENVDATPKVDKIEVVDAKTLNVTLSDGTKETVTLEK  242

Qy   240  KVAPNK--DLTVKVKNQSFVTKFVYEV-----------KKLAVE-KLTFDDDRAGQA  282
              |||:  |:  ||: |  :::||                :|:      :|:|::::|:|
Db   243  ALEPNKETEVTFKKIKDVEYKCAKVTYVVTATAVKSVSATNLKEVVVEFDGTVDKETABDA  302

Qy   283  IAFKLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIFEGGEATSTTGKL------AVG  336
              ::          : |:|:|        ||:|           |  |     |||
Db   303  ANYALKSGK------------TIKSVSLAADNKTATV------TLTDKLNNNKADAIS  342

Qy   337  I---KQGDYKVEV------------QVTKRGGLTVSNTGIITVK-NLDTPASAIKNVVF  379
              :   |:| |:|             |:|:   |:|:   |:   :||:|||: ||:  :
Db   343  ISNVKAGDKEINVKNVEFTAVDNKIPEVTE-----VKSLGTKAVKVTLSEPVENLSSTNF  397

Qy   380  ALD--------ADNDGVV--NYGSK----------LSG-KDFA----LNSGN--LVV  409
              :|:           :       | |                |:      ||       |:|
Db   398  TLDGKAYFGNVVMGAGNKTVILTPYSSSALSVGDHKLTVSGAKDFAGFVSLNSTHEPKVV  457

Qy   410  GEK---------ASLNKLVATIAGEDKVVDPGSISIKSSNHGIIS------------VV  447
              ||:            : :    |:  ||  |   | :
Db   458  EDKEAPTVTEATATLBTVTLTFS-EDIDMD---TVKASNVYWKSGDSKKEASEFERIAD  512

Qy   448  NNYITAEAAGEATL-TIKVGDVTKDVK----FKVTTDSRKLVSV--KANPD--KLQVVQ  497
              ||  |:||||| :| |||:||:||:      ||:| ||||:    : ::       |:|:
Db   513  NKYKFVFKGSEKTLPFGKVDVYVEDIKDYSDNKIAKDTKVTVTPEIDQTRPEVRKVTALD  572

Qy   498  NKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVVTTDSGSIGTKTIGVTGN  557
              ||:    |  |:  ::::        :::| :    |:     |::::::  ::    : |
Db   573  EKTIKVTPSKTVDGESAIKTGNYTVKDKDDK--VVS---VDKVTVDS------------  614

Qy   558  DVGEGTVHFQNGNGATLGSLY--VNVTEGNVAFKNFELVSKVGQVGQSPDTKLDLNVSTT  615
              ||  :  |  || |            |||     :   |      :|
Db   615  ---------KDSKSVIIDLYSKVSVGENTITIKNVKDATKLN-----NTMLD------  652

Qy   616  VEYQLSKYTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVDI------  669
              :   : | :|:  ::::      :       |:| :| : ::::||||:
Db   653  ---------YTGKFTRSDKEGPD-YE-----HVINADAK--AKKVVL--KPDKKMDAASLADY  696

Qy   670  --HLTKNGATAGKAATVEIVQETIAIKSV-NFKPVQTENFVEK----------------  708
              ::|      :  :    :     ::|:|  :|:  :    |:|
Db   697  SNYLVKI-----NDTLQTLSEDVATLSVSNDATVVTITFAETIIKGDDVVPASGKAISGSG  751

Qy   709  KINIGTVLELEKSNLDDIVKGINLTKETQHKVRVKSGAEQGKLYL----------DR  756
              |:|:         |||:|:||||:|:        |   | |        ||
Db   752  KVNVN---ELQ----VMGVKDTSGNVHK----KFNGSENKITLSSSTSTPLKLLAKIDK  797

Qy   757  NGDAVFNAG--DVKLGDVTVSQTSDSALPN-FKADLY--DTLTTKYTDKGTLVFKVLKDK  811
              ||    :|:  | |:|||| :|::::||   ||::|   |:| ||: ||::|: |:|:||
Db   798  DYDAKYTAELVDRKITVKVKFSTVINSAAANAFTSESHKIDSIQVNGTSTVTVKF-----K  852

Qy   812  DVITSEIGSQAVHVNV  827

Db   853  DEININASDLDLKVNL  868
```

RESULT 6
ID SLAP_ACEKI STANDARD; PRT; 762 AA.
AC P22258;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell surface protein precursor (S-layer protein).
OS Acetogenium kivui.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=2325;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=DSM 2030;
RX MEDLINE=90036724; PubMed=2681162;
RA Peters J., Peters M., Lottspeich F., Baumeister W.;
RT "S-layer protein gene of Acetogenium kivui: cloning and expression in
RT Escherichia coli and determination of the nucleotide sequence.";
RL J. Bacteriol. 171:6307-6315(1989).
RN [2]
RP PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
RC STRAIN=DSM 2030;
RX MEDLINE=92281680; PubMed=1596358;
RA Peters J., Rudolf S., Oschkinat H., Mengele R., Sumper M.,
RA Kellermann J., Lottspeich F., Baumeister W.;
RT "Evidence for tyrosine-linked glycosaminoglycan in a bacterial
RT surface protein.";
RL Biol. Chem. Hoppe-Seyler 373:171-176(1992).
RN [3]
RP DOMAINS.
RX MEDLINE=94156823; PubMed=8113161;
RA Lupas A., Engelhardt H., Peters J., Santarius U., Volker S.,
RA Baumeister W.;
RT "Domain structure of the Acetogenium kivui surface layer revealed by
RT electron crystallography and sequence analysis.";
RL J. Bacteriol. 176:1224-1233(1994).
CC -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
CC   of proteins which coat the surface of bacteria.
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC   layer with hexagonal symmetry.
CC -!- PTM: THE CARBOHYDRATE CONTENT OF THIS PROTEIN IS ABOUT 8% WHICH
CC   CORRESPOND TO ABOUT 40 TO 50 SUGAR MOLECULES PER MONOMER. O-LINKED
CC   GLYCANS CONSIST OF GLC, GALNAC AND GLCNAC.
CC -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; M31069; AAA21930.1; -.
DR PIR; A34355; A34355.
DR InterPro; IPR001119; SLH.
DR Pfam; PF00395; SLH; 2.
DR PROSITE; PS01072; SLH_DOMAIN; 2.
KW Signal; Glycoprotein; Repeat; Cell wall; S-layer.
FT SIGNAL 1 26
FT CHAIN 27 762 CELL SURFACE PROTEIN.
FT DOMAIN 30 94 SLH 1.
FT DOMAIN 95 145 SLH 2.
FT DOMAIN 146 204 SLH 3.
FT DOMAIN 473 479 SER/THR-RICH.
FT DOMAIN 625 630 SER/THR-RICH.
FT CARBOHYD 297 297 O-LINKED (GLC. . .).
FT CARBOHYD 516 516 O-LINKED (GLC. . .).
FT CARBOHYD 520 520 O-LINKED (GLC. . .).
FT CARBOHYD 632 632 O-LINKED (GLC. . .).

```
SQ   SEQUENCE   762 AA;  82785 MW;  34E9C784DECA67E CRC64;

Query Match
Best Local Similarity   6.3%;  Score 265.5;  Pred. No. 0.0001;  DB 1;  Length 762;
Matches  193;  Conservative  116;  Mismatches  347;  Indels  193;  Gaps  44;

Qy     47  AVIFTKILNLPVDENAQPSFKDAKNIWSSKYIAAVEK---AGVVKGDGKENFYPEGKIDR  103
           |::||. .   ::.       :. . ::  ::: .:.    .|::||:||||||||||||
Db     15  ALVFSA---MAVGFAATTPFTDVKD-DAPYASAVARLYALNITNGVGDPKFGVDQPVTR   69

Qy    104  A---SFASMLVSAYNLKDKVNGELVTTFEDL-LDHWGEEKANILINLGISVGTG-GKWEP  158
           |   .:::: |..|.: . :. :   :.  : ::|||||||||||||:||||| ||  :
Db     70  AQMITFVNRMLGYEDLAEMAKSE-KSAFKDVPQNHWAVGQINLAYKLGLAQGVGNGKFDP  128

Qy    159  NKSVSRAEAAQFI--Ai----TDKKYGKKDNAQAYVTDVKVSEPTKLTLTGTGLDKLSAD  212
           |.: ::|: | ::  .:    ||||||||||||||||
Db    129  NSELRYAQALAFVLRALGFKDLDWPYGYLAKAQ---DLGLVHGLNLAYNG---------  175

Qy    213  DVTLEGDKAVAIEASTDGTSAVVTLGGKVAPNKDLTVKVKNQSFVTKFVTEVKKLAVEK-  271
           .:.: || : .:. :.  :. .::.||||:.:||||||||||||||.|||.|.|||.
Db    176  -VIKRGDLALIIDRALE-VPMVKYVDGKEVLGEPLISKVATKAEYTVIATNAQDRSVEEG  233

Qy    272  --LTFDDDRAGQAIAFKLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIFEGGE--AT  327
             :.:     .|..: .:: ..|. .:  :  : .|.|.: :.  .. . :||   .|
Db    234  KVAVLDKDGKLITINAGLVDFS----EYLG----KVIVYSERFGDPVYVAEGDNDVVS   284

Qy    328  STTGKLAVG---IKQGDYKVEVQV-------------TKRGLTVSNTGIITVKNLDTPAS  372
           .| |. .:|   .|:.|.   ::.             || |.|  :  :.|.. .: .
Db    285  FTEGQDSVGTTVYKNDDNKTAIKVDDNAYVLYNGYLTKVSKVTVKEGAEVTIIN------  338

Qy    373  AIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLNKLVATIAGEDKVVDPG  432
           .| |..  :.:.::  |:.   : |.|  :|..| :. |.  |.||. |  :. ||.:
Db    339  --NNYLIVNGSYDNSTIVYNDVQSGDKYLNRDSNY-----ELKGTVTVTGAVSKVTD--  388

Qy    433  SISIKSSNH-------------GIISVVNNYITABAAGEATLTIKVGDVTK----DVK   473
           .|.||.::              ||:.:  : :     .| ||:|:.:..     :|:
Db    389  --IKANDYIYYGKQYDVDVNGNVVGTVIYVRNQVT---GTVTEKSVSGSTYKASIDNVS  441

Qy    474  FKVTTDS-------RKLVSVKANPDKLQV--VQNKTLPVTFVTTDQYGDPFGANTAAIK  523
           :.|.|:.       :|.:.  |. .:::.  ..:.:|.|:   |.  .|||  |..|.|
Db    442  YTVADNNVWNQLEPGKKVTVILNKDNVIVGISSTTTTTAVNYAIFKEKSDPFTAWFAKVK  501

Qy    524  EVLPKTGVVABGGLDVVTTD-------SGSIGTKIIGVTG--NDVGEGTVHFQNGNGAT  573
           .:|| .   :   ||::  :       :|.| | ::.: .  :..  . :.::.:  :
Db    502  LIlPD---AAEKVFDAVYSDVYDKVNLAEGTIVTYTVVDANGKLNDI------QRANDQP  551

Qy    574  LGSLYVN------VTEGNVAFKNFELVSKVGVGYGQSPDTKLDLNVSTTVEYQLSKYTSDRV  628
           : ||:.:      ||.   .|:|  .:.  | ||.|   |||:|:| : :  .  :.
Db    552  FSSASYKADAKVLTEGSTTY-----------YITDNTVLLN-NTSDGYKALKLTD----  594

Qy    629  YSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKT-------PGKVDIHLTKNGATAGKA  681
           .  . |  .:| .||||.|||.|:|.||.|.   :       ||||||||||||.||.:
Db    595  LKDATNL-NVKIVADNYNVAKVVFNNASFVSTTTSVYAYVTGTADVYV--NGSTFTRL   651

Qy    682  TVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLELEKSN--LDDIVKGINLTKETQHK  739
           ||.:.:  :|: :|.  .| ..:  .||:|:|.|:|.|:.:  |||:|||:|||||.:.:
Db    652  T---VLENGQTKTVDANAQLATNVTHKAV-VLTLTNAKIANIALPTVASGVKLTNIDQAN  707

Qy    740  VRVVKSGAEQGKLYLDRNGDAVFNAGDVK-LGDVTVSQTSDSALPNFKADLYDTLTTKYT  798
           :|:.||.|:.||||||:||||||||||:| |||||||||||||||||||||||||||||:
Db    708  LRITDT-TNKGYL-LDPNFIVDTNGNLKGLSDIT-----------KDTGVNLYT       749

Qy    799  DKGTLVFKV  807
           :|||:|||:
Db    750  NDVGKVFVI  758
```

```
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE   (Alkaline cellulase).
OS   Bacillus sp. (strain KSM-635).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1415;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91037937; PubMed=2230718;
RA   Ozaki K., Shikata S., Kawai S., Ito S., Okamoto K.;
RT   "Molecular cloning and nucleotide sequence of a gene for alkaline
RT   cellulase from Bacillus sp. KSM-635.";
RL   J. Gen. Microbiol. 136:1327-1334(1990).
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC       linkages in cellulose, lichenin and cereal beta-D-glucans.
CC   -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC       hydrolases)
CC   -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; M27420; AAA22304.1; -.
DR   PIR; S29043; S29043.
DR   PDB; 1G01; 31-DEC-02.
DR   PDB; 1G0C; 31-DEC-02.
DR   InterPro; IPR005086; CBM_17_28.
DR   InterPro; IPR008979; Gal_bind_like.
DR   InterPro; IPR015547; Glyco_hydro_5.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF03424; CBM_17_28; 2.
DR   Pfam; PF00150; cellulase; 1.
DR   Pfam; PF00395; SLH; 3.
DR   PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR   PROSITE; PS01072; SLH_DOMAIN; 2.
KW   Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat;
KW   3D-structure.
FT   SIGNAL        1     29
FT   CHAIN        30    941       ENDOGLUCANASE.
FT   DOMAIN       40     99       SLH 1.
FT   DOMAIN      100    151       SLH 2.
FT   DOMAIN      152    225       SLH 3.
FT   ACT_SITE    373    373       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    485    485       NUCLEOPHILE (BY SIMILARITY).
SQ   SEQUENCE   941 AA;  104628 MW;  BEA2AC3B169BFADA CRC64;

Query Match
Best Local Similarity   6.0%;  Score 252.5;  Pred. No. 0.0005;  DB 1;  Length 941;
Matches  193;  Conservative  109;  Mismatches  291;  Indels  311;  Gaps  42;

Qy      5  FPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQP   64
           |    . |.  ::|:.:.  :..|| .|. .. .:|:| :::|||.||:|:|:|||:|.:|
Db     41  FSDVKKTSWSFPYIKDLYEQEVITGTSAITFSPTDSVTRAQFTVMLTRGLGLEASSKDYP  100

Qy     65  SFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNL--------  116
           .|||:.| |: : |.|.| ||:|||.. |:| |.   : : .|.:..:|
Db    101  -FKDRKN-WAYKEIQAAYEAGIVTGKTNGEFAPNENITREQMAAMAVRAYEYLENELSLP  158

Qy    117  --KDKVNGELVTTF-EDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIA  172
             ::: .:.:.|:| .  ..  .::|. :|.|:  .  : :..:  :.   : .:::|
Db    159  EEQREYNDSSISTFAQDAV----QKAYVLELM-EGNTDGYFQPKRNSTREQSAKVIS   211

Qy    173  LTDKKYGKKDNAQAYVTDVKVSEPTK-----LTLTGTGLDKLSADDVTLEGDKAVAIE-  225
           :|. ..  ::|.||.|  :||| :       |:|:| |: |:|.  :| ||.| . :
Db    212  TLLWKVASHD--YLVHTEA-VKSPSEAGALQLVELNG--------QLTLAGEDGTPVQL  259

Qy    226  --ASTDGT--SAVVTLGGKVAPNKD------LTVKVKNQSFVTK-----FVVEVKKLA  268
             .|:|.|   |.|.|.||:|.|:|       |||:.|||.||:       |||||:|.
```

RESULT 7
GUN_BACS6
```
ID   GUN_BACS6     STANDARD;     PRT;     941 AA.
AC   P19424;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
```

```
Db  260 RGMSTHGLQWFGEIVNENAFVALSNDWGSNMIRLAMYIGENGYATNPEVKDLVYEGIELA 319
                                                        :||  |  | : |
Qy  269 VEKLTF--------DDDRAG-------QAIAFKLNDEKGNADVEYL--NLANHDVK     307

Db  320 FEHDMYIVDWHVHAPGDPRADVVSGAYDFFEEIADHYKDHPKN---HYIIWELANEP--  374

Qy  308 FVANNLDGSP--ANIPEGGEATSTTGKLAVGI--KQGDYKVEV---QVTKRGGLTVSNTG 360
                    :  ::                            :||
Db  375 --SPNNNGGPGLTNDEKGWEAVKEYAEPIVEMLREKGDNMILVGNPNWSQRPDLSADNP- 431

Qy  361 IITVQNL------------DTPAS-----AIKNVVFALD------------           382
                 : |:             |||  :|
Db  432 -IDAENIMYSVHFYTGSHGASHIGYPEGTPSSERSNVMANVRYALDNGVAVFATEWGTSQ 490

Qy  383 ADNDG------VVNYGSK--LSGKDPALNSQNLVVG--EKASLNKLVATIAGEDKVV    429
        |||:         : ::: : | :
Db  491 ANGDGGPYFDEADVWLNFLNKHNISWANWSLTNKNEISGAFTPPELGRTDAT------DL 544

Qy  430 DPGSISIKSSNHGIISVVNNYITAEAAGEATLTIKVGDVTKDVK------------FKVT 477
        |||:                    : :.:                          :|
Db  545 DPGANQVWAPEE--LSLSGEYVRARIKGIEYTPI--DRTKFFTKLVWDFNDGTTQGFQVN 599

Qy  478 TDS--RKLVSVKANPDKLQV----VQN------------                      498
         ||                     |||
Db  600 GDSPNKESITLSNNNDALQIEGLNVSNDISEGNYWDNVRLSADGWSENVDILGATELTID 659

Qy  499 ------------KTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVV               532
                    :  |||
Db  660 VIVEEPTTVSIAAIPQGPAAGWANPTRAIKVTEDDFESFGDGYKALVTITSEDSPSLETI 719

Qy  533 AEGGLD-------VVTTDSGSIGTKTIGVTGNDVGEGTVHFQNGNGATLGSLY-----   578
                    : : |:::  :  ||
Db  720 ATSPEDNTMSNIILFVGTEDADVISLDNITVSGTEIEIVIHDEKGT-ATLPSTFEDGTR  778

Qy  579 ------------VNVTEGN-------------VAFKNFELVSKVGQ               599
                    :   |||
Db  779 QGWDWHTESGVKTALTIEBANGSNALSWEYAYPEVKPSDGWATAPRLDFWKDELVRGTSD 838

Qy  600 Y------------GQSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEG           637
        |                |||||
Db  839 YISFPDFYIDAVRASEGAISINAVFQPPANGYWQEVPTTFEIDLTELDSATVTSD--ELYH 896

Qy  638 YEVE                                                         641
        |||:
Db  897 YEVK                                                         900
```

RESULT 8

```
ID   YDEK_ECOLI     STANDARD;      PRT;   1325 AA.
AC   P32051; P76140; P77168;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical lipoprotein ydek precursor (ORFT).
GN   YDEK OR ORFT OR B1510.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97251357; PubMed=9097039;
RA   Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA   Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA   Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA   Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA   Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA   Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT   "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 28.0-40.1 min region on the linkage map.";
RL   DNA Res. 3:363-377(1996).
RN   [3]
RP   SEQUENCE OF 595-1325 FROM N.A.
RX   MEDLINE=94100243; PubMed=8274505;
RA   Cartwright P.J., Timms M.W., Lithgow T., Hoej P.B., Hoogenraad N.J.;
RT   "An Escherichia coli gene showing a potential ancestral relationship
RT   to the genes for the mitochondrial import site proteins ISP42 and
RT   MOM38.";
RL   Biochim. Biophys. Acta 1153:345-347(1993).
-!-  SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
     (Potential)
-!-  SIMILARITY: TO E.COLI YFAL.
-!-  SIMILARITY: SOME, TO FUNGAL MITOCHONDRIAL IMPORT SITE PROTEINS
     ISP42 AND MOM38.
-!-  CAUTION: Ref.3 sequence differs from that shown due to a
     frameshift in position 653.
CC   -----------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------
DR   EMBL; AE000248; AAC74583.1; -.
DR   EMBL; D90793; BAA15190.1; ALT_INIT.
DR   EMBL; D90794; BAA15197.1; ALT_INIT.
DR   EMBL; X73295; CAA51730.1; ALT_FRAME.
DR   PIR; A64905; A64905.
DR   EcoGene; EG11780; ydeK.
DR   InterPro; IPR000437; Prok.lipoprot_S.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Hypothetical protein; Membrane; Lipoprotein; Signal;
KW   Complete proteome; Palmitate.
FT   SIGNAL      1     18       POTENTIAL.
FT   CHAIN      19   1325       HYPOTHETICAL LIPOPROTEIN YDEK.
FT   LIPID      19     19       N-palmitoyl cysteine (Potential).
FT   LIPID      19     19       S-diacylglycerol cysteine (Potential).
FT   CONFLICT  884    884       N -> K (IN REF. 3).
FT   CONFLICT 1317   1317       M -> S (IN REF. 3).
SQ   SEQUENCE   1325 AA;  136514 MW;  26A3A066FA19AD7D CRC64;

Query Match            5.5%;  Score 230;  DB 1;  Length 1325;
Best Local Similarity  19.2%;  Pred. No. 0.0073;
Matches  210;  Conservative 154;  Mismatches 367;  Indels 364;  Gaps 54;

Qy   11 GHWAEGSINYLVDKGAITGK------------PDGTYGPTESIDRASAAVIFTKILNLPVDENAQ   63
                                         :: ||.:
Db  296 GYFGNGTVN-ISNNGLINNKEYSLVGVQDGSHG------------VVN--VTDKGH           336

Qy   64 PSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASML--VSAYNLKDKVN       121
             :|.:                                        :.:
Db  337 WNFLGTGEAFRYIYIGDA------------GDGELNVSSEGKVDSGIITAGMKETGTGNITVKDK  389

Qy  122 GELVTTFEDLLDHWGEEKANI----NLG------ISVGTGGKWEPNKSVS               163
               :|:                            .:
Db  390 NSVITNLGTNLGYDGHGEWNISNQGLVVSNGGSSLGYGETGVGNVSITTGGMWEVNKNV-      448

Qy  164 RABAAQFIALTDKKYGKKDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVA      223
             |||  |     ||||          |||   ||||      |||||
Db  449 ----------------YTTIGVAGVGNLNISDGG--KFVSQNITFLGDKASG              482

Qy  224 IEAST--DGTSAVTLGGKVAPNKDLTVKVKNQSFVTKFVYEVKKLAVEKLTFDDDRAGQ      281
```

```
RESULT 9
YDBA_ECOLI

ID   YDBA_ECOLI     STANDARD;      PRT;   2003 AA.
AC   P33666; P76087; P76088; P76856; P76857; P76859;
DT   01-FEB-1994 (Rel. 28, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein ydbA.
GN   YDBA OR B1401/B1405.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97251357; PubMed=9097039;
RA   Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA   Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA   Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA   Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA   Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA   Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT   "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 28.0-40.1 min region on the linkage map.";
RL   DNA Res. 3:363-377(1996).
RN   [3]
RP   SEQUENCE OF 464-2003 FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=92190038; PubMed=1665988;
RA   Moszer I., Glaser P., Danchin A.;
RT   "Multiple IS insertion sequences near the replication terminus in
RT   Escherichia coli K-12.";
RL   Biochimie 73:1361-1374(1991).
CC   -!- SIMILARITY: TO S.TYPHIMURIUM ORF NEAR CYSG (AC P25928).
CC   -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR
CC       THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT
CC       BETWEEN AMINO ACIDS 839 AND 840.
CC   ------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------
DR   EMBL; AE000237; AAC74483.1; ALT_SEQ.
DR   EMBL; AE000237; AAC74487.1; ALT_SEQ.
DR   EMBL; D90778; BAA15009.1; ALT_SEQ.
DR   EMBL; D90778; BAA18880.1; ALT_SEQ.
DR   EMBL; D90779; BAA18881.1; ALT_SEQ.
DR   EMBL; X62680; -; NOT_ANNOTATED_CDS.
DR   Ecogene; EG11307; ydbA.
KW   Hypothetical protein; Complete proteome.
FT   CONFLICT     489    489    I -> V (IN REF. 2).
FT   CONFLICT     495    495    I -> V (IN REF. 2).
SQ   SEQUENCE   2003 AA;  205949 MW;  B83A12C8B53220BE CRC64;
```

```
Query Match             5.3%;  Score 223.5;  DB 1;  Length 2003;
Best Local Similarity   22.6%;  Pred. No. 0.024;
Matches  173;  Conservative  96;  Mismatches  275;  Indels  223;  Gaps  38;
```

```
Qy   115 NLKDKVNGELVTTFEDLLDHWGEEKANILINL-GISVGTGKWEPKSVSRAEAAQFIAL 173
             :|:||.|   .|  :::          |:|:|       |::|
Db   173 SLQDS-NGRKAT-----INLWQIDEANNTVALEGVSADGATKWQYNHN--------- 214

Qy   174 TDKKYGKKDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSA 233
                                                :|:|
Db   215 -----------------------------------GELVITG-----DNATVNNNGKTTVDGKDSTGTE 243

Qy   234 VVTLGGKVAPNKDLTVKKNQSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEKGN 293
             :  | |
Db   244 INGNNGKVIQDGDLDVSGGHGI--------DITGDSATV---DNKGT 280

Qy   294 ADVEYLNLANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGG 353
```

```
Db   483 IGTLNLMDATSFDTVGINVGNFGSGIVNVSNGATLNSTGYGF--------IGGNASGK 533
             |||.| .|.|
Qy   282 AIAFKLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIFEGGEATSTTG--------KL 333

Db   534 GIVNISTDSLWN------------LKTSSTNAQLQVGVLGTGELNITTGGIVKARDTQI 581

Qy   334 AVGIK-QGDYKVE----------VQVTKRGGLTVSNTGIITVKNLDT-----PASAI 374
             |.| |||       :|.|||   :|| ::|||
Db   582 ALNDKSKGDVRVDGQNSLLETFNMYVGTSGTGTLTLTNNGTLNVEGGEVVLGVFEPAVGT 641

Qy   375 KNV-----------------------VFALDADNDGVVNY-------GSKLSGK-- 398
             |.|
Db   642 LNIGAAHGEAAADAGFITNATKVEFGLGEGVFVFNHTNNSDAGYQVDMLITGDDKDGKVI 701

Qy   399 ---DFALNSQNLVVGGEKASLNKLVATIAGE--DKVVDPGISIKSSNHGIISVVNNYIT 452
                 :||    |||
Db   702 HDAGHTVFNAGNTYSG-KTLVNDGLLTIASHTADGVTGMGSSEVTIANPGTLDIL---AS 757

Qy   453 ABAAGRATLT-------IKVGDVTKDVKFKVT--------- 477
                :||| |||
Db   758 TNSAGDYTLTNALKGDGLMRVQLSSSDKMFGFTHATGTEFAGVAQLKDSTFTLERDNTAA 817

Qy   478 -------TDSRKLJVSVKANPDKL-----QVVQNKTLPVT-----FVTTD- 509
                :||:
Db   818 LTHAMLQSDSENTTSVKVGEQSIGGLAMNGGTIIFDTDIPAATLAEGYISVDTLVVGAGD 877

Qy   510 ----------QYGDPFGANTAAIKEVLPKTGVVAEGLDVVTTDS--G 545
                        :|:  | ||  :| :|:
Db   878 YTWKGRNYQVNGTGDVLIDVPKPWNDPMANNPLTTLNLLEHDD--SHVGVQLVKAQTVIG 935

Qy   546 SIGTKTI-GVTGNDV-GEGTVHF-QNGNGATLGSLYVNVT--EGNVAPKNF------- 591
             :|||||
Db   936 SGGSLITLRDLQGDEVEADKTLHIAQNGTVVAEGDYGFRLTTAPGNGLYVNVGLKALNIHG 995

Qy   592 ----ELVSKVGGYGQSPDTKL------DLNVSTTVEYQLSKYTSD------RVYSD 631
                    :
Db   996 GQKLTLAEHGGAYGATADMSAKIGGSGDLAINTVRQVSLSNGQNDYQGATYVQMGTLRTD 1055

Qy   632 PENLEGYEVESKNLAVADAKIV---GNKVVV----TGKTPGKV-----DIHLTKNGATAGK 680
             :  :|||
Db   1056 ADGALG---NTRELNISNAAIVDLNGSTQTVTFTGMGSTVLFKEGALTVNKGGISQGE 1112

Qy   681 AT-----VEIVQETIAIKSVNFK-PVQTENFVEKKINIGTVLELEKSNL--DDIVKGINLT 733
             :|:|:|
Db   1113 LTGGGNLNVTGGTLAIEBGLNARYNALTSISPNAEVSLDNTGLGRGNIANDGLLTLKNVT 1172

Qy   734 KETQHKVR---VVKSGAEQ-----------GKLYLDRNGDAVF------NAGDVKL---G 770
                :  :
Db   1173 GELRNSISGKGIVSATARTDVELDGDNSRFVGQFNID-TGSALSVNEQKNLGDASVINNG 1231

Qy   771 DVTVSQTSDSALPNFKADLYDTLTTKYTDKGTLVFKVLKDK------DVITSEI-GSQ 821
             |:||
Db   1232 LLTISTERRSWAMTHSISGSGDV-----TKLGTGILTLNNDSAAYQGTTDIVGGEIAFGSD 1286

Qy   822 AV------HVNVLNN 830
             :|:|
Db   1287 SAINMASCHINIHNS 1301
```

```
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF00395; SLH; 2.
DR   PROSITE; PS01072; SLH_DOMAIN; 2.
KW   Signal; Cell wall; S-layer; Repeat.
FT   SIGNAL        1     53
FT   CHAIN        54   1116       SURFACE LAYER PROTEIN.
FT   DOMAIN       57    120       SLH 1.
FT   DOMAIN      121    171       SLH 2.
FT   DOMAIN      172    231       SLH 3.
SQ   SEQUENCE   1116 AA;  123397 MW;  86D583D7AC72546F CRC64;

Query Match             5.2%;  Score 219;  DB 1;  Length 1116;
Best Local Similarity  19.2%;  Pred. No. 0.018;
Matches  202;  Conservative 142;  Mismatches 347;  Indels 362;  Gaps 47;

Qy   15 EGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQ------PSFKDA  69
Db   68 EKTVKRLEALGLVAGYGNGDFGADKITRAEFATLIVRARGL--EQGAKLAQFNTTYTDV 125

Qy   70 KNI-WSSKYIAAVEKAGVVKGDKGENFYPEGKIDRASFASMLVSAYNLKDKVNGELVTTF 128
Db  126 RSTDWFAGFVNVASGEEIVKGFPDKSFKPQNQVTYAEAVTMIVRALGYEPSVRGV-- 180

Qy  129 EDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYGKKDNAQ-- 185
Db  181 ------W----PNSMISKGSELNIA------KGINNPNMQQPAATIFKML--DNALRVK 221

Qy  186 ----AYVTDVK--VSEPTKLT-LITGTGLDKLS 210
Db  222 LMEQIEYGTDIRLNVTDETLLTKYLKVTVRDMDWAHEKGNNSDELPLVTNVPAIGLGSLK 281

Qy  211 ADDVTLEGDKA---VAIEASTDGTSAVVTLGGK----VAP 243
Db  282 ANEVTLNGKDADLGSNTYKVAEGINPNAFDGQKVQVWIKDDRENVIVWMEGSEDEDVVM 341

Qy  244 NKDLTVKVKGNQSFVTKFYVYEVKKLAVEKLTFDDDRAGGAIAFKLNDE--- KGNAD 295
Db  342 DRVSALYLKGKAFTDDIVKDLSKSDLDDVKIEMD--GSEKSYRLTEDTKITYNFTRFNDP 399

Qy  296 VEYLNLANHD----VKFVAN-NLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVT 349
Db  400 VDALSKIYKDNDIFGVKVVVLNDNNEVAYLHIID---DQTIDKSVKGVKYG 446

Qy  350 KRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNY-GSKLS----GKDFA 401
Db  447 ------SKVISKIDADKKKITNLDNSKFSDLEDQDEGKDFL 481

Qy  402 LNSQNLVVGEKASLNKLVATIAGEDKVVDPGSISIKSSNHGIISVVNNYITABAAGEATL 461
Db  482 V-----FLDGQPAKL-------GDLKESDVYSVYYADGDKDKYLVFANRNVAEGKVEKVV 529

Qy  462 T-------IKVGDVT--------KDVKFKVTTDSRKLVSVKANPDKLQVVQNKT 500
Db  530 SRNKTDIRLTVGGKTYKVYPDASYSENANKDVK-KVNSDLLIISNLDGBEVKLLL------583

Qy  501 LPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVVTTDSGSIGTKTIGVTGN- 557
Db  584 ------DPSG----RVRHIETKDAIDDRKPLAIITK-----GATYNSSK 617

Qy  558 DVGEGTVHFQNGNGATLG---SLY----VNVTBGNVAPKNFE--LVSKVGQYGQSPDTK 607
Db  618 DTYDFTVMTQKGKTQIVSLDQKDIYDRYGVNYDKSNDKRQAFEKDLVELLQPKVKEDSA 677

Qy  608 LDLNVSTTVEYQL-------SKYTSDRVYSDPENLEG-YEVESKNLAV 647
Db  678 TDANQTVLLEVNFDSKGEVDKVKVLDSKLKYSEKSTWDKLADEDDDVVGDYEVTDK--T 734

Qy  648 ADAKIVGNKVVVTGKTPGKVDIHLTKNGATA------------GKATVEIV-- 686
Db  735 AVFKMTGDLTPATGTKRGEL----KNAGTAKFKDVAKKSDLKVWYSVDEDKGEVGEVQAIFV 789

Qy  687 ---------------QETIAIKSVNFKPVQTENF------VEKKINIG 713
```

```
Db  281 MTVT--------------DPESMGIQDGDKAIVNNEGESTITNGGTGTGTQINGDD------------ 321
Qy  354 LTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGS-KLSGKDFALNSQNLVVGEK 412
Db  322 ATANNNGKTTVDGKDSTGTEI------NGNNGKVIQDGDLDVSGGGHGID----ITGDS 370
Qy  413 ASLNKLVATIAGEDKVVDPGSISIKSSNHGIISVVNNYITABAAGEATLT-------IKV 465
Db  371 ATVDN-----KGTMTVTDPESIGIQVD--GDQAVVNN------EGESAITNGGTGTGTQING 417

Qy  466 GDVTKDVKFKVTTDSRKLVSVKANPDKLQVVQNKTLPVT------FVTTDQYG-DPFGAN 518
Db  418 DDATANNNGKTTVDGKDSTGTEIAGNNGKVIQDGDLDVSGGGHGIDITGDSATVDNKGTM 477

Qy  519 TAAIKEVLPKTGVVAEGGLDVV-----TTDSGSIGTKTIG--VTGNDVGEGTVHFQNGN 570
Db  478 TVTDPESI---GIQIDGDQAIVNNEGESTITNGGTGTGTQINGNDATANNSGKTTVDGKDST 534

Qy  571 GATL-------GSLYY-----NVTE-GNVAFKNFELVSKVGQY-------GQSP 604
Db  535 GTKIAGNIGIVNLDGSLTVTGGAHGVENIGDNGTVNNKGDIVVSDTGSIGVLINGEGATV 594

Qy  605 DTKLDLNVS-------TTVEYQLSKYTSDRV--YSDPENLEGYEVESKNLAVADAKIVG 654
Db  595 SNTGDVNVSNEATGFSITTNSGKVSLAGSMQVGDFSTGVDLNGNN-NSVTLAAKDLKVVG 653

Qy  655 -------NKVVVTG-------KTPGK------------VDIHLTKNGATA-GKATV 683
Db  654 QKATGINVSGDANTVNITGNVLVDKDKTADNAAEYFFDPSVGINVYGSDNNVTLDGKLTV 713

Qy  684 EIVQEFIAIKSVNF--KPVQTENFV---EKKINIGTVLEL--EKSNLDDIVKGINLTKE 735
Db  714 VSDSEVTSRQSNLFDGSAEKTSGLVVIGDGNTVNMNGGLELIGEKNALADGSQVTSLRTG 773

Qy  736 TQHKRVRVKSGABQGKLYLDRNGDAVFNAGDVKLGDVTVSQTSDSAL 782
Db  774 YSYTSVIVVSG--ESSVYL--NGDTTI-SGEFPLGFAGVIRVQDKAL 815
```

```
RESULT 10
SLPH_BRECH
ID   SLPH_BRECH   STANDARD;   PRT;   1116 AA.
AC   P38538;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Surface layer protein precursor (Hexagonal wall protein) (HWP).
OS   Brevibacillus choshinensis.
OC   Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
OX   NCBI_TaxID=54911;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 54-68.
RC   STRAIN=HPD31;
RX   MEDLINE=90170842; PubMed=2307650;
RA   Ebisu S., Tsuboi A., Takagi H., Naruse Y., Yamagata H., Tsukagoshi N.,
RA   Udaka S.;
RT   "Conserved structures of cell wall protein genes among protein-
RT   producing Bacillus brevis strains.";
RL   J. Bacteriol. 172:1312-1320(1990).
CC   -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC       layer with hexagonal symmetry.
CC   -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D90050; BAA14103.1; -.
```

```
Db   790  VDGSGLGGDHQFGMVKQYGTASKQDTITIVTKDGDSVTEKEYKLDGDADDLKVDQDIRRG  849
Qy   714  TVLELEKSN-----LDDIVKGIN-----------------LTKETQHKVRVVKSGAEQGKLYLD  755
           :  :|.: :|  |:: :|:::
Db   850  DVISFTLNSDGEVIVDDDVVEVNNNHIDNTASKSATLMPEDERQ----KAGID--KLVVA  903
Qy   756  RNGDAVFNAGDVKLGDVTVSQTSDSALPNFKADLYDTLT-TKYTDKG-----------  801
           :|: :|.::
Db   904  RVDEVDGNTISLNYADGKTQKYYTKASTAF-IDVIDGLEGIDGDVDEGDYIVMIDSADIDG  962
Qy   802  ---TLVFKVLKDKDVITSEIGSQAVHVNVLNNP  831
             |:.:|:
Db   963  TRFDYVLVVSSDDEIRTQHISTKAV-TDFLNKP  994
```

RESULT 11
OMPB_RICCN
ID   OMPB_RICCN   STANDARD;   PRT;   1655 AA.
AC   Q9KKA3; Q9KK98; Q9XC45;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Outer membrane protein B precursor (168 kDa surface-layer protein)
DE   (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rOmpB)
DE   (rOmp B) (Contains: 120 kDa surface-exposed protein (Surface protein
DE   antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
GN   OMPB OR RC1085.
OS   Rickettsia conorii.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=781;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Malish 7;
RX   MEDLINE=21442074; PubMed=11557893;
RA   Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA   Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA   Raoult D.;
RT   "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL   Science 293:2093-2098(2001).
RN   [2]
RP   SEQUENCE OF 33-1649 FROM N.A.
RC   STRAIN=Indian tick typhus, and Malish 7;
RX   MEDLINE=20393643; PubMed=10939649;
RA   Roux V., Raoult D.;
RT   "Phylogenetic analysis of members of the genus Rickettsia using the
RT   gene coding the outer-membrane protein rOmpB (ompB).";
RL   Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
RN   [3]
RP   SEQUENCE OF 353-1655 FROM N.A.
RC   STRAIN=Malish 7;
RA   Stenos J., Walker D.;
RT   "The rickettsial outer membrane protein A and B genes of Rickettsia
RT   australis, the most divergent rickettsia of the spotted fever group.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC   STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC   VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC   similarity).
CC   -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC   (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC   layer with hexagonal symmetry (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.

DR   EMBL; AE008659; AAL03623.1; -.
DR   EMBL; AF123721; AAF34124.1; -.
DR   EMBL; AF123726; AAF34129.1; -.
DR   EMBL; AF149110; AAD39533.1; -.
DR   PIR; E97835; E97835.
DR   InterPro; IPR006315; Autotransport.
DR   InterPro; IPR005546; Autotransporter.
DR   Pfam; PF03797; Autotransporter; 1.
DR   TIGRFAMs; TIGR01414; autotrans_barl; 2.
KW   Antigen; S-layer; Cell wall; Complete proteome.
FT   CHAIN   1   1334   120 kDa SURFACE-EXPOSED PROTEIN.
FT   CHAIN   1335   1655   32 kDa BETA PEPTIDE.
FT   VARIANT   61   61   P -> A (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT   75   75   G -> S (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT   78   78   K -> N (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT   251   251   V -> A (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT   413   413   N -> D (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT   959   959   I -> V (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT   988   988   A -> T (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT   1139   1139   R -> L (IN STRAIN INDIAN TICK TYPHUS).
FT   CONFLICT   353   354   KD -> GH (IN REF. 3).
FT   CONFLICT   776   776   F -> S (IN REF. 3).
FT   CONFLICT   1159   1159   B -> D (IN REF. 3).
FT   CONFLICT   1177   1177   G -> D (IN REF. 3).
FT   CONFLICT   1492   1492   H -> R (IN REF. 3).
SQ   SEQUENCE   1655 AA; 168342 MW; E49E19377D5FCE37 CRC64;

Query Match   5.2%;
Best Local Similarity   20.5%;   Score 217; DB 1; Length 1655;
Matches 183; Conservative 123; Mismatches 336; Indels 252; Gaps 45;

Qy   104 ASFASMLVSAYNLKDKVNGELVTTFBDLLDHWGEEKANILINLGISVTGGKWEPNKSVS 163
           |||| : :: ::   : :: : ::: ::   ::
Db   26 ASFAGSAMGAAIQNRTTNAVATTVDGVGFDQTAVPANVAVPLNAVITAG----VNKGIT   81

Qy   164 -RAEAAQPIAL--TDKKYG---KKDNAQAYVTDV------------KVSEPTKLTLTGTGLD 207
           : : : ::: |:| :::|
Db   82 LNTPAGSFNGLFLNTANNLDVTVREDTTLGFTNVVNNANHFNLMLNAGKTLTTTGQGIT 141

Qy   208 KL-----SADDVTLEGDKAVAIE------------------------AST-------DG 230
            | ::::|
Db   142 NVQAAATKNANNVVAQVNNGAAIDNNDLQGVGRIDCGAAASTLVFNLANPTTQKAPLIG 201

Qy   231 TSAVVTLGGKVAPN-KDLTVKVKNQSFVTKFVEVKKLAVEKLTFDDDRAGQAIAFKLND 289
           :|:   :: :
Db   202 DNAVIVNGANGTLNVTNGFIKVSSKSPAT-------------VNVINIGD--GQGIMFNTDA 248

Qy   290 EKGNADVEYINLANHDVKFVANNLDGSPANIFEGGEATST----TGKLAVGIKQ----GD 341
            : : :
Db   249 D---NVNTINLQANGATITFNGTDGTGRLVLLSKNAAATDFNVTGSLGGNLKGIIEFNT 304

Qy   342 YKVEVQVTKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFA 401
           ::| :: :   :
Db   305 VAVNGQLKANAG---ANAAVIGTNN-----GAGRAAGFVVSVDNGKVATIDGQVYAKDMV 356

Qy   402 LNSQNLV--------VGE------KASLNKLVATIAGEDKVVDPGSISIK---------- 437
           |:::|::   :::
Db   357 IQSANAVGQVNFRHIVDVGTDGTTAFKTAASKVAITQNSNFGTTDFGNLAAQIIVPNTMT 416

Qy   438 --------SSNHG-----------------IISVNNYITABAAGEATLTIKVGDV 468
                                             :: : :
Db   417 LNGNFTGDASNPGNTAGVITFDANGTLASASADNVAVTNNITAIEASGAGVVQLS-GTH 475

Qy   469 TKDVK----FKVTTDSRKLVSVKANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAA 521
           ::::   :: :
Db   476 AAELRLGNAGSVFKLADGT--VINGKVNQTAL------------------VGGALAAGTIT 516

Qy   522 IKEVLPKTGVVAEGG---LDVVTTDSGSIGTKTIGVTGNDVGE--GTVHFQNGNGATLG 575
            :|::   :|: :: ::   ::
Db   517 LDGSATITGDIGNAGGAAALQGITLANDATKTLTLG-GANIIGANGGTINFQ-ANG---G 571

Qy   576 SLYVNVTEGNVAFKNFELVSKVGQYGQSPPTKLDLNVSTTVEYCLSKYTSDRVYSDPENL 635
           :: ::| : :|:|
```

```
Db  572 TIKLTSTQNNIVV-DFDLAIATDQTGVVDASSLITNAQTLTINGKIG-----TVGANNKTL  625
                     |:  |  ||||    ||
Qy  636 EGYEVESKNLAVADAKIVGNKVVVTGKIPGKVD------IHLTKNGATAGAKATVEIV--   686
                 :::|  ::|:           |::|:||
Db  626 GQFNIGSSKTVLSDGDVAINELVIGNN-GAVQFAHNTYLITRTTNAAGQGKIIFNPVVN    683
Qy  687 QETIAIKSVNF---KPVQTENFVEK-KINIGTVLELEKSNLDDIVKGINL--TKETQHK    739
                            |  :||:||      :|:||
Db  684 NNTTLATGTNLGSATNPLAEINFGSKGAANVDTVL------NVGKGVNLYATNITTTD    735
Qy  740 VRVVK---------SGAEQGKLY----LDRN-----GDAVFNAGDVKLGDVTVS       775
                           ||   || :     ||::   ||  | ::|| |||:
Db  736 ANVGSFIFNAGGTNIVSGTVGGQQGNKFNTVALDNGT-TVKFLGNATFN------GNTTIA  789
Qy  776 QTSDSAL-PNFKADLYDTLTTKYTDKGTLVFKVLKDKDVITSEIGSQAVHVNVL        828
               : :| :|  :|             ||    ||  ||  :: |  ||
Db  790 ANSTIQIGGNYTADFVASA------DGTGIVEFV-NTGPITVTLNKQAAPVNAL        836
```

RESULT 12
SLPM_BACBR
ID   SLPM_BACBR   STANDARD;   PRT;   1053 AA.
AC   P06546;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Middle cell wall protein precursor (MWP).
OS   Bacillus brevis (Brevibacillus brevis).
OC   Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
OX   NCBI_TaxID=1393;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=47;
RX   MEDLINE=88115203; PubMed=2828336;
RA   Tsuboi A., Uchihi R., Adachi T., Sasaki T., Hayakawa S., Yamagata H.,
RA   Tsukagoshi N., Udaka S.;
RT   "Characterization of the genes for the hexagonally arranged surface
RT   layer proteins in protein-producing Bacillus brevis 47: complete
RT   nucleotide sequence of the middle wall protein gene.";
RL   J. Bacteriol. 170:935-945(1988).
RN   [2]
RP   SEQUENCE OF 1-199 FROM N.A.
RC   STRAIN=47;
RX   MEDLINE=87137282; PubMed=3029027;
RA   Yamagata H., Adachi T., Tsuboi A., Takao M., Sasaki T.,
RA   Tsukagoshi N., Udaka S.;
RT   "Cloning and characterization of the 5' region of the cell wall
RT   protein gene operon in Bacillus brevis 47.";
RL   J. Bacteriol. 169:1239-1245(1987).
RN   [3]
RP   SEQUENCE OF 676-1053 FROM N.A.
RC   STRAIN=47;
RX   MEDLINE=87008404; PubMed=2428810;
RA   Tsuboi A., Uchihi R., Tabata R., Takahashi Y., Hashiba H., Sasaki T.,
RA   Yamagata H., Tsukagoshi N., Udaka S.;
RT   "Characterization of the genes coding for two major cell wall
RT   proteins from protein-producing Bacillus brevis 47: complete
RT   nucleotide sequence of the outer wall protein gene.";
RL   J. Bacteriol. 168:365-373(1986).
RN   [4]
RP   SEQUENCE OF 1-50 FROM N.A.
RC   STRAIN=47;
RX   MEDLINE=90078123; PubMed=2512285;
RA   Tsuboi A., Uchihi R., Engelhardt H., Hattori H., Shimizu S.,
RA   Tsukagoshi N., Udaka S.;
RT   "In vitro reconstitution of a hexagonal array with a surface layer
RT   protein synthesized by Bacillus subtilis harboring the surface layer
RT   protein gene from Bacillus brevis 47.";
RL   J. Bacteriol. 171:6747-6752(1989).
CC   -!- FUNCTION: The middle wall protein binds to peptidoglycan and to
CC   the outer cell wall protein.
CC   -!- SUBUNIT: THE MIDDLE CELL WALL LAYER IS COMPOSED OF SUBUNITS OF
CC   THE MIDDLE CELL WALL PROTEIN. THESE PROTEINS FORM A HEXAGONAL
CC   ARRAY WITH A LATTICE CONSTANT OF 14.5 NM IN THE MIDDLE CELL
CC   WALL LAYERS.
CC   -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC   layer with hexagonal symmetry.
CC   -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC
DR   EMBL; M15364; AAA87321.1; -.
DR   EMBL; M14238; AAA22372.1; -.
DR   EMBL; M19115; AAA22760.1; -.
DR   EMBL; M31828; AAA22619.1; -.
DR   PIR; A28555; A28555.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF00395; SLH; 2.
DR   PROSITE; PS01072; SLH_DOMAIN; 2.
KW   Cell wall; S-layer; Signal; Repeat.
FT   SIGNAL   1   23
FT   CHAIN   24   1053   MIDDLE CELL WALL PROTEIN.
FT   DOMAIN   29   92   SLH 1.
FT   DOMAIN   93   143   SLH 2.
FT   DOMAIN   144   203   SLH 3.
SQ   SEQUENCE   1053 AA; 117146 MW; DB4213188D9D5B4F CRC64;

Query Match   5.0%; Score 211; DB 1; Length 1053;
Best Local Similarity   19.9%; Pred. No. 0.036;
Matches 193; Conservative 138; Mismatches 321; Indels 320; Gaps 45;

```
Qy   15 EGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQ--PSFKDA      69
            ::        |||          ::            |||
Db   40 EKTVKRLEALGLVAGYGNGEYGVDKTITRAEFATLVVRARGL--EQGAKLAQFSNTYTDV   97
Qy   70 KNI-WSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGELVTTF   128
Db   98 KSTDWFAGFVNVASGEIVKGFPDKSFKPQNQTYAEAVTMIVRALGYEPSVKGV-----    152
Qy  129 EDLLDHWGB---EKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYGKKDNA-  184
                          ||:|||
Db  153 ------WPNSMISKASEL-NIARSITT-----PNNAATRGDIFKML------DNAL      190
Qy  185 ------QAYVIDVK--VSEPTKLT-------LTGTGLD                        207
                                  ::  ::|:
Db  191 RVDLMEQVBFGTDIRHEITKETLLTKYLKVTVRDMEWAQEAGNDSEDLPLVTNVPAIGLG  250
Qy  208 KLSADDVTLEGDKAVAIEAST----DGTSAVVTLGGKVA--PNKDLTVKYKNQSFVT     258
            |:||||  ||||        :|:      |||::
Db  251 KIKANEVTLNG-KDAGIGNTTYKVADGINANDFDGQVQVWIKDDKEDVIVMEGSTQE     309
Qy  259 KFVEVKKLAVEKLTFDDDRAGQAIAFKINDEKGNADVEYLNLANHDVKFVANNLDGSPA   318
Db  310 VIMDRVGEFTLKGKTFEDPK-------DLSNSDL-------                     336
Qy  319 NIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLTVSNTGI-ITVKNLDTPASAIKNV  377
            :|::                      :::    ||:
Db  337 ------ADLKLELDASEKSYRFNKNTKVTYNFTRFNDPVDGLKEI                 375
Qy  378 VFALDADN-DGVVNYGSKLSGKDFALNSONLV---VGEKASLNKLVATIAGEDKVVDPG   432
            :  |      ||                :||           ||  :||
Db  376 I----KDNADGGFTFGAKV-----VLDNNNEIAYIHVIDDQSMNK------EDEGVKYG   419
Qy  433 SISIKSSNHGIISVVNNYITAEAAGEATLTIKVGDVTKDVKFKVTTDSRKLVSVKANPDK  492
Db  420 S------------------------EVISIKDTDKKKITN--RDNDK               440
Qy  493 LQVVQNKTLPVTFVTTDQYGDPFGANTAAIKE----------VLPKTGVVAEGG        536
            :        :                                       | ||
Db  441 FNDLDGKEEGKDFLVF-LNGKP--AKFSDLKEGMVYSVYYADGDEDKLLVFAIDTVVEGK  497
```

```
CC      VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC   -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC   -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC       layer with hexagonal symmetry.
CC   -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC       ----------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC       ----------------------------------------------------------------
DR   EMBL; L04661; AAB48987.1; -.
DR   PIR; JN0896; JN0896.
DR   InterPro; IPR006315; Autotransport.
DR   InterPro; IPR005546; Autotransporter.
DR   Pfam; PF03797; Autotransporter; 1.
DR   TIGRFAMs; TIGR01414; autotrans_barl; 1.
KW   Antigen; S-layer; Transmembrane; Cell wall.
FT   CHAIN         1    1353       120 kDa SURFACE-EXPOSED PROTEIN.
FT   CHAIN      1354    1645       32 kDa BETA PEPTIDE.
FT   TRANSMEM   1415    1429       MEMBRANE ANCHOR (POTENTIAL).
FT   CONFLICT    657     657       H -> N (IN REF. 2).
FT   CONFLICT    842     842       V -> I (IN REF. 2).
FT   CONFLICT   1071    1071       G -> A (IN REF. 2).
FT   CONFLICT   1306    1306       G -> S (IN REF. 2).
SQ   SEQUENCE   1645 AA;  169698 MW;  0CB5541C7EB185EE CRC64;

Query Match             4.9%;  Score 207.5;  DB 1;  Length 1645;
Best Local Similarity   20.6%;  Pred. No. 0.091;
Matches  184;  Conservative 130;  Mismatches 346;  Indels 235;  Gaps  47;

Qy      104 ASFASMLVSAYNLKDKVNGELVTTPEDLLDHWGEEKANILNLGISVGTGGKWEPNKSVS 163
                 : :       || ||  ::         ::
Db       26 AGFSGVAMGAVMQYNRTTNAAATT---VDGAGFDQTGAGVNLPVATNSVITANSNNAIT 81

Qy      164 -----------RABAAQFIALTDKKYGKKDNAQAYVTDV---------KVSEPTKLTLTG 203
                        | :|||:|:|||
Db       82 FNTPNGNLNSLFLDTANTLAVTI---NENTTLGFVTNVTKQGNFFNFTIGAGKSLTITG 137

Qy      204 TGL------DKLSADDVTLEGDKAVAIE-------ASTDGTS-------------AV 234
                                    : |:
Db      138 HGITAQQAATTKSAQNVVSKVNAGAAINDNLSGVGSIDFTAAPSVLEFNLINPTTQEAP 197

Qy      235 VTLG--GKVAPNK----DLT---VKVKNQSFVTKFVYEVKKLAVEKLITFDDDRAGQAIAF 285
                  :                          ::          GIKTINIGDN---QGLMF
Db      198 LTLGDNAKIVNGANGILNITNGFVKVSDKTFA------------------------- 244

Qy      286 KLNDEKGNADVEYLNL--ANHDVKFVANNLDGSPANIFEGGEAT--STTGKLAVGIKQGD 341
                                  :|
Db      245 NTTPDAANA---LNLQGGGNTINFNGRDGTGKLVLIVSKNGNATEFNVIGSLGGNLKG-- 298

Qy      342 YKVEVQVTKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFA 401
                :                                             ::  :
Db      299 -VIEFDTTAAAGKLIANGGAANAV-IGTDNGAGRAAGFIVSVDNGNAATISGQVVAKDIV 356

Qy      402 LNSQNLVVGKASLNKLVATIAG-----------EDKVDPGSISKSSNHGIISVV----N 448
                :|:         |:|                   |:|
Db      357 IQSAN--AGGQVTFEHLVDVGLGGKTNFKTADSKVIITENASFGSTDFGNLAVQIVVPNN 414

Qy      449 NYITABAAGEATLTIKVGDVTKDVKFKVTTDSRKLVSVKANPDKLQVVQNKTLPVTFVTT 508
                :|:                |:        ::  :   :|:
Db      415 KILTGNFIGDAK--NNGNTAGVITFNA---NGTIVSGNTDPN-IVVTNKAIEVEGAGI 467

Qy      509 DQYGDPFGA-----NTAAIKEVLPKT---------GVVAEGGLDVVTTDS 544
                :|:
Db      468 VQLSGIHGAELRLGNAGSIFKLADGTVINGPVNQNPLVNNALAAGSIQLDGSAIITGDI 527

Qy      545 G----------------SIGTKTIGVTG-NDVGB---GTVHFQNGNGATLGSLYNVTEGNV 586
                                 : | :    ::   :   :     :  ::   :   |:
Db      528 GNGAVNAALQDITLANDASKILTLSGANIIGANAGGAIHFQ-ANG---GTIQLTSTQNNI 583
```

```
Qy      537 LDVVTTDSGSIGTKTIGVTGNDVGEGTVHFQNGNGATLG---------------SLYV 579
                :|:                |||       :|
Db      498 VDKVVSRNNNDYRLITIGDKTYRVYEGATFSDDGNKDVQDIDKDHWDLVDSLDDETVKLYL 557

Qy      580 NVTEGNVAF---------KNFELVSKVGQYGQSPDTKLDLNVST------TVE------ 617
                 :|:            |||                       |:|
Db      558 DAS-GRVRHIETKDAIDDRKQKAIVTRSATFNTSKDT-WDFRVLTQKGKEITVSLEAKNI 615

Qy      618 YQLSKYTSDRYYSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVD-IHLTKNGA 676
                  :                   | ::   | :    |         :
Db      616 YDFDGKNFSRDNKNQDDLEDIIVPSKD---KDTLLLEVTLDADGK-PGKVEFLKPVKVEQ 671

Qy      677 TAGKA---------------------TVEIVQETIAI---------KSVNFKPVQTENF 705
                ||:|                     :| :  |:               |:
Db      672 BSGKAWDDLADEDDDMVGDYEVTDKTAVFNMTGKLEESSKRKELKNAKTAKFKDVADEND 731

Qy      706 VEKKINIGTVLELEKSNLDDIVKGINLTKETQHKVRVRVKSGABQQKLYL---DRNGDAVF 762
                  :|::            |: |||              :|:
Db      732 LSV---IYTVNDKDEVEAIFVVEGDGLTGDA-HYGQVIDFGRKGGKDTIRVWEKDGDKVV 787

Qy      763 NAGDVKL-GDVTVSQTSDSALPNFKADLYDTLTTKYTD--------KGTLV---FKV 807
                  :|:              | ::   :    :|
Db      788 EK-EYKLDGDQDDLKDEDIRRNDFIAFTVDSNDEVVVDDVVEVVNKNAKGMLAEVTDEKG 846

Qy      808 LKDKDVITSEIG 819
                :||::   : :
Db      847 MKDANIDKMVVG 858
```

RESULT 13
```
OMPB_RICTY
ID   OMPB_RICTY   STANDARD;    PRT;   1645 AA.
AC   P96989;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Outer membrane protein B precursor (168 kDa surface-layer protein)
DE   (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rOmpB)
DE   (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE   antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN   OMPB OR SLP.
OS   Rickettsia typhi.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=785;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wilmington;
RX   MEDLINE=94040787; PubMed=8224886;
RA   Hahn M.-J., Kim E.-K., Kim I., Chang W.-H.;
RT   "Cloning and sequence analysis of the gene encoding the crystalline
RT   surface layer protein of Rickettsia typhi.";
RL   Gene 133:129-133(1993).
RN   [2]
RP   PARTIAL SEQUENCE.
RC   STRAIN=Wilmington;
RX   MEDLINE=92114896; PubMed=1370573;
RA   Ching W.M., Carl M., Dasch G.A.;
RT   "Mapping of monoclonal antibody binding sites on CNBr fragments of
RT   the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT   prowazekii.";
RL   Mol. Immunol. 29:95-105(1992).
RN   [3]
RP   IDENTIFICATION OF CLEAVAGE SITE.
RC   STRAIN=Wilmington;
RX   MEDLINE=92104668; PubMed=1729180;
RA   Hackstadt T., Messer R., Cieplak W., Peacock M.G.;
RT   "Evidence for proteolytic cleavage of the 120-kilodalton outer
RT   membrane protein of rickettsiae: identification of an avirulent
RT   mutant deficient in processing.";
RL   Infect. Immun. 60:159-165(1992).
CC   -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC       STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
```

```
Qy   587 AFKNFELVSKVGQYQGSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEGYEVESKNLA 646
              :|:|                                 : : :|| |   ||
Db   584 -LVDFPLDVTTDQTGVDASSLTNNQTLINGSIG-----TIGANTKTLGRFNVGSSKTI 637

Qy   647 VADAKIVGNKVVVTGKTPGKVDIHLTK---------NGATAGKATV--EIVQETIAIKS 694
             |:                        |         ||  ||            |
Db   638 LNAGDVAINEIVM--ENDGSV--HLTHNTYLITKTINAANQGKIIVAADPINTDIALADG 693

Qy   695 VNF----KPVGTENFVEKKINIGTVLELEKSNLDDIVKGINL----TKETQHKV------ 740
           ::     |                 |:|:                 ||
Db   694 TNLGSAESPLSNIHFATKAANGDSIH--------IGKGVNLYANNITTTDANVGSLHPR 745

Qy   741 ---RVVKS--GAEQG----KLYLDRN------GDAVFNAGDVKLGDVTVSQTSDSALPN 784
              ::|      |:       |:  |       |     |  |::||
Db   746 SGGTSIVSGTVGGQQLKLNNLILDNGTTVKFLGDITFNGG-TKIEGKSILQISSNYITD 804

Qy   785 F--KADLIYDTLTTKYTDKGTLV-------RKVLKD------KDVITSEIGSQAVH 824
          :   |   |  |  |         |:|          |:| ::| |
Db   805 HIESADNTGTLEFVNTDPITVTLNKQGAYFGVLKQVMVSGPGNIAFNEIGNGVAH 859
```

RESULT 14

YEEJ_ECOLI

```
ID   YEEJ_ECOLI     STANDARD;      PRT;   2358 AA.
AC   P76347; P94750;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical protein yeeJ.
GN   YEEJ OR B1978.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III., Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97251358; PubMed=9097040;
RA   Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA   Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA   Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA   Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,
RA   Saito N., Sampei G., Seki Y., Sivasundaram S., Tagami H.,
RA   Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
RT   "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 40.1-50.0 min region on the linkage map.";
RL   DNA Res. 3:379-392(1996).
CC   -!- SIMILARITY: Contains 13 Big-1 domains.
CC   -!- SIMILARITY: Belongs to the intimin/invasin family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE000289; AAC75042.1; ALT_INIT.
DR   EMBL; D90837; BAA15800.1; -.
DR   EMBL; D90836; BAA15799.1; ALT_INIT.
DR   EcoGene; EG13378; YeeJ.
```

```
DR   InterPro; IPR003344; Big_1.
DR   InterPro; IPR003535; Intimin.
DR   InterPro; IPR009964; Invasin_intimin.
DR   InterPro; IPR002482; LysM.
DR   InterPro; IPR000601; PKD.
DR   Pfam; PF02369; Big_1; 13.
DR   PRINTS; PR01169; INTIMIN.
DR   SMART; SM00634; BID_1; 13.
DR   SMART; SM00257; LysM; 1.
DR   SMART; SM00089; PKD; 6.
KW   Hypothetical protein; Repeat; Complete proteome.
FT   DOMAIN      738    834    BIG-1 1.
FT   DOMAIN      840    931    BIG-1 2.
FT   DOMAIN      932   1033    BIG-1 3.
FT   DOMAIN     1042   1137    BIG-1 4.
FT   DOMAIN     1146   1237    BIG-1 5.
FT   DOMAIN     1246   1350    BIG-1 6.
FT   DOMAIN     1351   1448    BIG-1 7.
FT   DOMAIN     1449   1553    BIG-1 8.
FT   DOMAIN     1554   1655    BIG-1 9.
FT   DOMAIN     1661   1754    BIG-1 10.
FT   DOMAIN     1763   1853    BIG-1 11.
FT   DOMAIN     1855   1950    BIG-1 12.
FT   DOMAIN     1952   2053    BIG-1 13.
FT   CONFLICT    105    105    S -> G (IN REF. 2).
SQ   SEQUENCE   2358 AA;  248599 MW;  23224975DBF631ED CRC64;
```

```
Query Match             4.9%;  Score 207.5;  DB 1;  Length 2358;
Best Local Similarity   20.2%;  Pred. No. 0.14;
Matches  188;  Conservative 126;  Mismatches 390;  Indels 225;  Gaps 39;
```

```
Qy    17 SINYLVDKGA---------ITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQPSFKDAKN 71
         :::::  ||                 :: ::|| ::||:|  : ||:| |: ::|
Db  1236 TVHFIGDTAAAKIIELAPVPDSIIAGTP--QNSSGSVITATV----VDNNGFPVKGVTVN 1289

Qy    72 IWSSKYIAAVEKAG--VVKGDGKENF------------------YPEGKIDRASFASMLVSA 113
             :  |::|  :                     :: ::::|  : :|  :::
Db  1290 FTSNAATAEMTNGGQAVTNEQGKATVTYNTRSSIESGARPDTVEASLENGS--STLSTS 1347

Qy   114 YN----------LKDKVN-GELVTTFEDLLDHWGEEKANILINLGISVGTG---- 153
          |:                : ::|  ||    :  :|| |: |:|:|
Db  1348 INVNADASTAHLTLLQALFDTVSAGETTSLYIEVKDNYGNGVPQQEVTLSVSPEGVTPS 1407

Qy   154 -----GKWEPNKSVSRAEAAQFIALTDKKYGKKDNAQAYVTDVKVSEPTKLTLTG 203
              ::::   :: |:  : ||: ::: |    ::::
Db  1408 NNAIYTTNHDGNFYASFTATKAGVYQLTATLB----NGDSMQQTVTYVPNVANAEITLAA 1463

Qy   204 TGLDKLSADDVTLEGDKAVAIEASTDG-----TSAVVTL-------------GGKVAPNKD 246
         |  :  :::|   :    |    :  |        ::|         ::
Db  1464 S-KDPVIADNNDL--TTLTATVADTEGNAIANTEVTFTTLPEDVKANFTLSDGGKVITDAE 1520

Qy   247 LTVKV-----------------------------KNQSFVTKFVVEVKKLAVEKLTFDDDRAGQAIAF- 285
         | ||:                           :  |  ||:::  |  |: :  :
Db  1521 GKAKVTLKGTKAGAHTVTASMTGKSEQLVVNFIADTLTAQVNLNVTEDNFIANNVGMTR 1580

Qy   286 ---KLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIF------EGGEA-TSTTGKLAV 335
              ::::::  :::|  |  |  :::|  ::|      |:::|  :: |:|
Db  1581 LQATVTDGNGNP--------------LANEAVTFTL-------PADVSASFTLGQGGSAITDINGKAEV 1628

Qy   336 ---GIKQGDYKVEVQVTKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYG 392
            :||:|| ||| |  |:  |:| :: : |  ||  |       ::: :  ::  :
Db  1629 TLSGTKSGTYPVTVSVNNYG---VSDT---------------------------KQVTLIADA-----GTAKLA 1667

Qy   393 SKLSGKDFALNSQ---------------------------NLVVGEKASLNKLVATIAGEDKVVDPGGSISIKS 438
         |:::  |  ||                          ::|  :  :| :|::|::
Db  1668 SLTSVVSFVVSTTEGAITMASVTDANGNPVEGIKVNFRGTSVTLS-----------STSVET 1718

Qy   439 SNHGIISVVNNYITAEAAGEATLITKVGD---------VTKDVKFKVTTDSRKLVSVKANPDKL 493
         ::  ::  ||  ::    :|  |  | |           ::| ::::  :  |:: :::
Db  1719 DDRGFAEIL---VTSEVGLKTVSASLADKPTEVISRLLNASADVNSATTISLEIPEGQV 1775

Qy   494 QVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVAEGGLDVVTTDSGSIGTKTIG 553
         ||:::::  ||:  |   :  : ||  |  :|:|||  ::: || ::  ||  |  :
```

```
Db  1776  MVAQD--VAVKAHVNDQFGNPVAHQPTFSAEPSSQMIISQN-----TVSTNTQGVAEVT  1828
             :|: .|    |       |  |   :  ||| :    |||
Qy   554  VTGNDVGEGTVHFQNGNGATIGSLYNVTEGNVAFKNFELVSKVGQYGQSPDTKLDLNVS  613

Db  1829  MTPERNGSYMVKASLPNGASLEKQLEAIDEKLTLTASSPLIGVYAPTGATLTATLTSANG  1888
           |:    |   : |.  |  | :.  |:    |      | |  | .   || || |.
Qy   614  TTVEYQLSKYTSDRVYSDPENLEGYBVESKNLAVADAKIVGNKV---VVTGKTPGKVDIH  670

Db  1889  TPVEGQVINFS--VTPEGATILSGGKVRTNSSGQAPVVLISNKVGTYTVTASFHNGVTIQ  1945
           | :| .| :    || |||   |||. | |.| |  |:| |||||||  ||||||||:|
Qy   671  -----LTKNGATAGAKATVEIVQETIAIKSVNFKPVQ------TENFVEKKINIGTVLEL  718

Db  1946  TQTTVKVTGNSSTAHVASFIADPSTIAATNTDLSTLKATVEDGSGNLIEGL----TVYFA  2001
            :.|  | ||   ||.|| |||||||.||.||:|||||||||||||||||||      :|
Qy   719  EKSNLDDIVKGINLTKETQHKRVRVKSGAEQGKLYLDRNGDAVFNAGDVKLGDVTV----  774

Db  2002  LKSGSATLTSLTAVTDQNGIATTSVK-GAMTGSVTV---SAVTTAGGMQTVDITLVAGP  2056
           |.||||||||||||||||||||||||  |||||||||    ||||||||||||||||||
Qy   775  SQTSDSALPNFKADLYDTLTTKYTDKGTL  803

Db  2057  ADTSQSVLKSNRS----SLKGDYTDSAEL  2081
```

RESULT 15
YEEJ_ECO57

```
ID   YEEJ_ECO57     STANDARD;      PRT;   2660 AA.
AC   Q8X8V7; Q8X2B9; Q8X2C0;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical protein yeeJ.
GN   Z3135 OR ECS2775/ECS2776.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11260655;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- SIMILARITY: Contains 16 Big-1 domains.
CC   -!- SIMILARITY: Belongs to the intimin/invasin family.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC       frameshift in position 1315.
```

```
DR   EMBL; AE005423; AAG57041.1; -.
DR   EMBL; AP002559; BAB36198.1; ALT_FRAME.
DR   EMBL; AP002559; BAB36199.1; ALT_FRAME.
DR   InterPro; IPR003344; Big_1.
DR   InterPro; IPR003535; Intimin.
DR   InterPro; IPR008964; Invasin_intimin.
DR   InterPro; IPR000601; PKD.
DR   Pfam; PF02369; Big_1, 16.
DR   PRINTS; PR01369; INTIMIN.
DR   SMART; SM00634; BID_1; 16.
DR   SMART; SM00089; PKD; 8.
KW   Hypothetical protein; Repeat; Complete proteome.
FT   DOMAIN    738    834    BIG-1 1.
FT   DOMAIN    840    929    BIG-1 2.
FT   DOMAIN    931   1033    BIG-1 3.
FT   DOMAIN   1042   1132    BIG-1 4.
FT   DOMAIN   1134   1236    BIG-1 5.
FT   DOMAIN   1245   1335    BIG-1 6.
FT   DOMAIN   1337   1439    BIG-1 7.
FT   DOMAIN   1448   1539    BIG-1 8.
FT   DOMAIN   1548   1652    BIG-1 9.
FT   DOMAIN   1653   1750    BIG-1 10.
FT   DOMAIN   1751   1855    BIG-1 11.
FT   DOMAIN   1856   1957    BIG-1 12.
FT   DOMAIN   1963   2056    BIG-1 13.
FT   DOMAIN   2065   2156    BIG-1 14.
FT   DOMAIN   2157   2252    BIG-1 15.
FT   DOMAIN   2254   2355    BIG-1 16.
SQ   SEQUENCE   2660 AA;  280062 MW;  01EB92A08F5C09D2 CRC64;
```

```
Query Match             4.9%;
Best Local Similarity   20.3%;
  Matches  197;  Conservative  134;  Mismatches  387;  Indels  253;  Gaps   46;

Score 207.5;  DB 1;  Length 2660;
Pred. No. 0.17;
Matches 197;  Conservative 134;  Mismatches 387;

Qy    17  SINYLVDKGA------ITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQPSFKDAKN   71
          :::::|::          |||||||||::|    ::      ||:      :   :
Db  1538  TVHFIGDTAAAKIIELTPVPDSIIAGTP--QNSSGSVITATV----VDNNGFPVKGVTVN  1591

Qy    72  IWSSKYIAAVBKAG--VVKGDGKENF------------------------YPEGKIDRASFASMLVSA  113
          |:          |:   :                                |:
Db  1592  FTSNAATAEMTNGGQAVTNEQGKATVTYNTRSSIESGARPDTVEASLENGS--STLSTS  1649

Qy   114  YNLKDKVNGELVTTFEDLLDHWGE-EKANTILNLGISVGTG-GKWEPNKSVSRAEAAQ--  169
          :|:  :     :| ||::::::|   ||| |||||:|:     ||  :|::|: :
Db  1650  INVNADASTAHILTLLQALFDTVSAGDTTNLYIEVKDNYGNGVPQQEVTLSVSPSEGVTPS  1709

Qy   170  ------------FIALTDKKYG---------KKDNAQAYVT-----------DV  191
                      ::||:|::|          |:||||||:            ::
Db  1710  NNAIYTNHDGNFYASFTATKAGVYQVTATLENGDSMQQTVTYVPNVANAEISLAASKDP  1769

Qy   192  KVSEPTKL-TLTGTGLD--------KLSADDVTL-------EGDKAVAIEASTDGTSA  233
          ::  :||| :|:|||:|        ||:|||||:       :|  |||:||:|||:||
Db  1770  VIANNNDLTTLTATVADTEGNAIANSEVTFLPEDVRANFTLGDGGKVV--TDTEG-KA  1825

Qy   234  VVTLGGKVAPNKDLTVKV--KNQSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAF-----  285
          ::||:|:    :|:||:|   |:||::|||:|||::|||:||||||||||:|:|
Db  1826  KVTLKGTKAGAHTVTRASMAGGKSEQLVVNFIADTLTAQVNLNVTEDNFIANNVGMTRLQA  1885

Qy   286  KLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIF------EGGEA-TSTTGKLAV--  335
          |:||:||||||: |:|||:|||||||||||||||||      :||:| ||:||||||
Db  1886  TVTDGNGNP--------LANEAVTFTL-----PADVSASFTLGQGGSAITDINGKAEVTLS  1933

Qy   336  GIKQGDYKVEVQVTKRGGLTVSNTGIITVKNLDTPASAIKNVFALDADNDGVVNYGSKL  395
          |:| | |:|:|:|::  ||||||||:||| |:::        |::|:||:|:|:|:|||
Db  1934  GTKSGTYPVTVSVNNYG--VSDT----------------------KQVTLIADA--GTAKLASLT  1972

Qy   396  SGKDFALNSQ-----------------NLVVGEKASLNKLVATIAGEDKVVDPGSISIKSSNH  441
          | :|:::                    |:::| |: :        :|     :|:::
Db  1973  SVVSFVVSTTEGATMTASVTDANGNPVEGIKVNFRGTSVTLS--------STSVETDDR  2023

Qy   442  GIISVNNYITAEAAGEATLITKVGD----VTKDVKFKVTTDSRKLVSVKANPDKLQVV  496
          | ::::   ::|  |:|| ::|:||    ::|     ::   |:::| || :: ::
Db  2024  GFAEIL--VTSTEVGLKTVSASLADKPTEVISRLLNAKADINSATTSLEIPEGQVMVA  2080
```

```
Qy    497 QNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVVTTDSGSIGTKTIGVTG 556
              |:|::|||:||||     :    :   ::: :|::|:   :
Db   2081 QD--VAVKAHVNDQFGNPIINESVTFSAEPPEHMTISQ---NIVSTDTHGIAEVTMTPER 2135

Qy    557 NDVGEGTVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQYGQSPDTKLDINVSTTV 616
              |:|||      :    |      ||      :   |:     :
Db   2136 N--GSYMVKASLANGSSYEKDLVIDQKLTLSASSPLIGVNSFTGATLTATLTSANGTPV 2193

Qy    617 EYQLSKYTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVDIHLTKNGA 676
              |:  |:  :   :|     ::    :    |    |||
Db   2194 EGQVINFS---VTPEGATLSGGKVRTNSSGQAPVVLTSNKV---GTYTVTASFH--NGV 2244

Qy    677 TAGKATVEIVQETIAIKSVNFKPVQTENFVEKKINIGTV---LELEKSNLDD----IVKG 729
                 |    :||      |::     :  ||::|
Db   2245 T-------IQTQTIVKVTGNSSTAHVASFIADPSTIAATNSDLSTLKATVEDGSGNLIEG 2297

Qy    730 INLTKETQHKVRVVKSGA---EQGKLYLDRNG---------------DAVFNAGDVKLG 770
               : :          |||||       |:|     ||    ::
Db   2298 LTV------YFALKSGSATLTSLTAVTDQNGIATTSVRGAITGSVTVSAVTTAGGMQTV 2350

Qy    771 DVTV----SQTSDSALPNFKADLYDTLTTKYTDKGTL---------VFKVLKDKDVITS 816
              |:|:       | |||   |:  :  |||      ||  || ::::
Db   2351 DITLVAGPADASQSVLKNNRS----SLKGDFTDSAELHLVLHDISGNPIKVSEGLEFVQS 2406

Qy    817 EIGSQAVHVNV 827
              |:|: |  ||
Db   2407 --GTNAPYVQV 2415
```

Search completed: April 7, 2004, 17:33:32
Job time : 24 secs

```
OM protein - protein search, using sw model

             GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

Run on:    April  7, 2004, 17:26:39 ; Search time 24 Seconds
                                  ; (without alignments)
                                  3338.648 Million cell updates/sec

Title:          US-09-844-281-1
Perfect score:  4202
Sequence:       1 AGKSFPDVPAGHWAEGSINY...........ITSEIGSQAVHVNVLNNPNL 833

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Database :      PIR_78:*
                1: pir1:*
                2: pir2:*
                3: pir3:*
                4: pir4:*

Searched:       283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:      283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

                         SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2833.5	67.4	874	2	JC4930	S-layer protein pr
2	719	17.1	814	2	I40048	S-layer protein pr
3	348	8.3	1176	2	A34355	surface-layer 125K
4	265.5	6.3	762	2	A34355	cell surface prote
5	255.5	6.1	1099	2	T14850	S-layer protein pr
6	252.5	6.0	941	2	S29043	cellulase (EC 3.2.
7	242.5	5.7	1036	2	T30311	S-layer protein -
8	240.5	5.7	4919	2	T31105	hypothetical prote
9	239.5	5.7	652	2	B59102	hypothetical prote
10	236	5.6	1109	2	A56143	surface-array prot
11	233.5	5.6	1386	2	AC1533	surface protein (L
12	231.5	5.5	2059	2	D82671	surface protein XF
13	231	5.5	1268	2	B99789	hemagglutinin/hemo
14	231	5.5	1270	2	B85649	hypothetical prote
15	230.5	5.5	2154	2	F83068	hypothetical prote
16	230	5.5	1325	2	A64905	ydeK protein - Esc
17	229	5.4	1612	2	AB1347	probable peptidogl
18	226.5	5.4	2020	2	C48399	ABC-type transport
19	225	5.4	1324	2	T18265	endo-1,3(4)-beta-g
20	225	5.4	3029	2	S76109	hypothetical prote
21	224	5.3	1806	2	AF1717	probable peptidogl
22	219	5.2	1116	2	A35129	surface layer prot
23	219	5.2	1343	2	E90893	hypothetical prote
24	218.5	5.2	1175	2	S70843	hemolytic protein
25	217	5.2	1483	2	C97012	probably celluloso
26	217	5.2	1655	2	E97835	hypothetical prote
27	217	5.2	5291	2	F90696	hypothetical prote
28	216.5	5.2	1536	2	A43855	high-molecular-wei
29	216	5.1	5188	2	B85547	probable RTX famil
30		215.5	5.1	1635	2	A70452	hemolysin/imporle
31		213.5	5.1	1461	2	E90696	hypothetical prote
32		213	5.1	1343	2	D85724	hypothetical prote
33		213	5.1	1477	2	B43855	high-molecular-wei
34		212.5	5.1	1461	2	A85547	hypothetical prote
35		212.5	5.0	1959	2	AG1085	middle wall protei
36		211	5.0	1084	2	A28555	large repetitive p
37		211	5.0	2256	2	AD1018	probable invasin Y
38		211	5.0	3013	2	AB0480	cell surface antig
39		209.5	5.0	2340	2	B71704	extracellular seri
40		209.5	5.0	2554	2	AB3528	hypothetical prote
41		209	5.0	404	2	F59097	hypothetical prote
42		209	5.0	694	2	AC2008	hypothetical prote
43		208.5	5.0	5627	2	C83339	hypothetical prote
44		207.5	4.9	1345	2	H90975	crystalline surfac
45		207.5	4.9	1645	2	JN0896	

```
                         ALIGNMENTS

RESULT 1
JC4930
S-layer protein precursor - Bacillus licheniformis
C;Species: Bacillus licheniformis
C;Date: 22-Oct-1996 #sequence_revision 01-Nov-1996 #text_change 15-Oct-1999
C;Accession: JC4930
R;Zhu, X.; McVeigh, R.R.; Malathi, P.; Ghosh, B.K.
Gene 173, 189-194, 1996
A;Title: The complete nucleotide sequence of the Bacillus licheniformis NM105 S-layer-e
A;Reference number: JC4930; MUID:97082965; PMID:8964497
A;Accession: JC4930
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-874 <ZHU>
A;Cross-references: GB:U38842; NID:g1055336; PIDN:AAC44405.1; PID:g1055337
A;Experimental source: strain NM105
C;Comment: This protein is a glycoprotein. It functions as protective coats, molecular
C;Genetics:
A;Gene: olpA
C;Superfamily: S-layer repeat homology
F;1-29/Domain: signal sequence #status predicted <SIG>
F;30-874/Product: S-layer protein #status predicted <MAT>
F;157-209/Domain: S-layer repeat homology <SLR3>

Query Match             67.4%;  Score 2833.5;  DB 2;  Length 874;
Best Local Similarity   67.3%;  Pred. No. 3.1e-125;
Matches  569;  Conservative  98;  Mismatches 156;  Indels  23;  Gaps   7;

Qy      1 AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     30 AGKSFPDVPAGHWAEDSINTLVDKGAIVGKPDGTYGPTESIDRASAAVIFTKILNLPVDE 89

Qy     61 NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV 120
            |||||||||||||||||||||||||||||||:||||||||||||||||||:|:||||||
Db     90 NAQPSFKDAKNLWSSKYIAAVEKAGVVKGDGKDNFYPEGKIDRASFASMLVGAYNLKEKV 149

Qy    121 NGELVTTFEDLLDHWGEEKANILINLGISVGTGKWEPNKSVSRAEAAQFIALTDKKYGK 180
            :|||:|:||||||||||||||||||:|:|||||||||||||||||||||||:||||||:
Db    150 DGTLVTKFDDLRGHWGEEKANILVNLGISVGTGKWEPNKSVSRAEAAAQFIALTKKYAK 209

Qy    181 KDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVTLGGK 240
            ||||||:||:|||||||||||||||||||||||||||||||||||||||||||||||||
Db    210 PENSDAKVTNVAATEPTQLTLTGTGLNKLTABDVTLEGNKAIALEASKDGKSAVVTLSGK 269

Qy    241 VAPNKDLTVKVKNQSFVTKFVYEVKCKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLN 300
            ||||||||||||:|||||:|:||:||||||||||||||||||||||||||||||:||||
Db    270 IAPNKELPVKVKGNTFIVKVYEVKKLRVEQLTFDDDRADQAVVFKLNDEKGNADIEYLD 329

Qy    301 LANHDVKFVANNLDGSPANLPEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLTVSNTG 360
            ||:||||||||||||:||||:||||:|||||||||||||:|||||||||||||:||||
Db    330 IAGHDVKFVANNLDGTPANLFBGGTAESTTGKLAVGIAEGKYKVEVQVTKRGGITVSNTG 389
```

```
A;Molecule type: DNA
A;Residues: 1-762 <PET>
A;Cross-references: GB:M31069; NID:g141844; PIDN:AAA21930.1; PID:g141845
C;Superfamily: S-layer repeat homology
F;32-85/Domain: S-layer repeat homology <SLR1>
F;96-150/Domain: S-layer repeat homology <SLR2>

Query Match              6.3%;  Score 265.5;  DB 2;  Length 762;
Best Local Similarity   22.7%;  Pred. No. 4.6e-05;
Matches  193; Conservative  116; Mismatches  347; Indels  193; Gaps  44;

Qy    47 AVIFTKILNLPVDENAQPSFKDAKNIWSSKYIAAVEK---AGVVKGDGKENFYPEGKIDR 103
         :::.:    :: .   .:
Db    15 ALVFSA---MAVGFAATTPFTDVKD--DAPYASAVARLYALNITNGVGDPKFGVDQPVTR  69

Qy   104 A---SFASMLVSAYNLKDKVNGELVTTFEDL-LDHWGEEKANILINLGISVTG-GKWEP 158
         |   :  :: :     ::
Db    70 AQMITFVNRMLGYEDLAEMAKSE-KSAFKDVPQNHWAVGQINLAYKLGLAQGVGNGKFDP 128

Qy   159 NKSVSRAEAAQPI--AL---TDKKYGKKDNAQAYVTDVKVSEPTKLTLTGTGLDKLSAD 212
         .:    :     : :       .  :
Db   129 NSELRYAQALAFVLRALGFKDLDWPYGYLAKAQ---DLGLVHGLNLAYNG--------- 175

Qy   213 DVTLEGDKAVAIEASTDGTSAVVTLGGKVAPNKDLTVKKNQSFVTKFVYKFVEVKKLAVEK- 271
         ::
Db   176 -VIKRGDIALILDRALB-VPMVKYVDGKEVLGEPLISKVATKAEYTVIATNAQDRSVEEG 233

Qy   272 -LTFDDDRAGQAIAFKLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIFEGGE-AT 327
         :
Db   234 KVAVLDKDGKLTTINAGLVDFS-------EYLG---KKVIVYSERFGDPVYVAEGDNDVVS 284

Qy   328 STTGKLAVG---IKQGDYKVEVQV------------TKRGGLTVSNTGIIITVKNLDTPAS 372
         :  . ::
Db   285 FTEGQDSVGTTVVKNDDNKTAIKVDDNAYVLYNGYLITKVSKVTVKEGAEVTIIN------- 338

Qy   373 AIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVGEKASLNKLVATIAGEDKVVDPG 432
         :   :
Db   339 --NNYLIVNGSYDNSTIVYNDVQSGDKYLNRDSNY------ELKGTVTVTGAVSKVTD-- 388

Qy   433 SISIKSSNH------------GIISVVNNYITAEAAGEATLTIKVGDVTK----DVK 473
         ::::.:
Db   389 ---IKANDYIYYGKQYDVNGNVVGTVIYVVRRNQVT----GTVTEKSVSGSTYKASIDNVS 441

Qy   474 FKVTTDS-------RKLVSVKANPDKLQV---VQNKTLPVTFVTTDQYGDPFGANTAAIK 523
         : :
Db   442 YTVADNNVWNQLEPGKKVTVILNKDNVIVGISSTTTTAVNYAIFKEKSDPFTAWFAKVK 501

Qy   524 EVLPKTGVVAEGGLDVVTTD-----SGSIGTKTIGVTG-NDVGEGTVHFQNGNAT 573
         :  ::
Db   502 LILPD---AAEKVFDAVVSDVYDKVNLAEGTIVTYTVDANGKLNDI------QRANDQP 551

Qy   574 LGSLYVN----VTEGNVAFKNFELVSKVGQYGQSPDTKLDLNVSTTVEYQLSKYTSDRV 628
         |:::: |
Db   552 FSSASYKADAKVLTEGSTTY--------------YITDNTVLLN-NTSDGYKALKLTD--- 594

Qy   629 YSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKT----------PGKVDIHLTKNGATAGKA 681
         :  :
Db   595 LKDATNL-NVKIVADNYNVAKVVVFNNASFVSTTTSTVYAYVTGADVVV--NGSTFTRL 651

Qy   682 TVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLELEKSN--LDDIVKGINLTKETQHK 739
         :
Db   652 T---VLENGQTKTYDANAQLATNYTHKAV-VLTLTNAKIANIALPTVASGVKLTNIDQAN 707

Qy   740 VRVVKSGAEQGKLYLDRNGDAVFNAGDVK-LGDVTVSQTSDSALPNFKADLYDTLTTKYT 798
         :
Db   708 LRITDT-TNKGYL-LDPNFIVVDTNGNLKGLSDIT---------------KDTGVNLYT 749

Qy   799 DKGTLVFKV 807
         :
Db   750 NDVGKVFVI 758
```

```
Qy   131 -LLDHWGEEKANILINLGISVGT--GGK-WEPNKSVSRABAAQFIALTDKKYGKKDNAQ 185
Db   157 STVKPWAKSYLEIAVANGVIKGSEANGKTNLNPNAPITRQDFAVVFSRT--------- 205

Qy   186 AYTDVKVVSEPTKLJLTGTGLDKLSADDVTLEGDKAVAIEAST----DGTSAVVTLGG 239
Db   206 --------------------IENVDATPKVDKIEVVDAKTLNVTLSDGTKETVTLEK 242

Qy   240 KVAPNK-DLTVKKNQSFVTKFVYEV----------KKLAVE-KLTFDDDRAGQA 282
Db   243 ALEPNKETEVTFKIKDVEYKAKVTYVVTTATAVKSVSATNLKEVVVEFDGTVDKETAEDA 302

Qy   283 IAFKLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIFEGGEATSTTGKL------AVG 336
Db   303 ANYALKSGK--------------TIKSVSLAADNKTATV-----TLTDKLNNNKADAIS 342

Qy   337 I---KQGDYKVEV------------QVTKRGGLTVSNTGIITVK-NLDTPASAIKNVVF 379
Db   343 ISNVYAGDKEINVKNVEFTAVDNKIPEVTE-----VKSLGTKAVKVTLSEPVENLSSTNF 397

Qy   380 ALD--------ADNDGVV--NYGSK------LSG-KDFA----LNSQN-LVV 409
Db   398 TLDGKAYFGNVVMGAGNKTVILTPYSSSALSVGDHKLTVSGAKDFAGFVSLNSTHEFKVV 457

Qy   410 GEK--------ASLNKLVATIAGEDKVVDPGSISIKSSNHGIIS--------------VV 447
Db   458 EDKEAPTVTEATATLETVTLFS-EDIDMD----TVKASNVYWKSGDSKKEASEFERIAD 512

Qy   448 NNYITAEAAGEATL-TIKVGDVTKDVK---FKVTTDSRKLVSV--KANPD--KLQVVQ 497
Db   513 NKYKFVFKGSEKTLPTGKVDVYVEDIKDYSDNKIAKDTKVTVTPEIDQTRPEVRKVTALD 572

Qy   498 NKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVVTTDSGSIGTKTIGVTGN 557
Db   573 EKTIKVTFSKTVDGSAIKTGNYTVKDKDDK--VVS---VDKVTVDS---------- 614

Qy   558 DVGEGTVHFQNGNGATLGSLY--VNVTEGNVAFKNFELVSKVGQYGQSPDTKLDLNVSTT 615
Db   615 --------KDSKSVIIDLYSKVSVGENTITIKNVKDATKLN------NTMLD------ 652

Qy   616 VEYQLSKYTSDRVISDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVDI------ 669
Db   653 ------YTGKFTRSDKEGPD-YE----HVINADAK--AKKVVL--KFDKKMDAASLADY 696

Qy   670 --HLTKNGATAGKATVEIVQETIAIKSV-NFKPVQTENFVEK----------- 708
Db   697 SNYLVKI-----NDTLQTLSEDVATLSVSNDATVVTITFAETIKGDDVVFASGKAISGSG 751

Qy   709 KINIGTVLELEKSNLDDIVKGINLTKETQHKVRVVKSGAEQGKLYL-----------DR 756
Db   752 KVNVN--ELQ----VMGVKDTSGNVHK----KFNGSENKITLSSTSTPLKLAKIDK 797

Qy   757 NGDAVFNAG--DVKLGDVTVSQTSDSALPN-FKADLY--DTLTTKYTDKGTLVFKVLKDK 811
Db   798 DYDAKYTAELVDRKTVKFSTVINSAAAANAFTSESHKIDSIQVNGTSTVTVKF----K 852

Qy   812 DVITSEIGSQAVHVNV 827
Db   853 DEINTNASDLDLKVNL 868
```

RESULT 4
A34355
cell surface protein precursor - Acetogenium kivui
C;Species: Acetogenium kivui
C;Date: 15-Jun-1990 #sequence_revision 15-Jun-1990 #text_change 08-Oct-1999
C;Accession: A34355
R;Peters, J.; Lottspeich, F.; Baumeister, W.
J;Bacteriol. 171, 6307-6315, 1989
A;Title: S-layer protein gene of Acetogenium kivui: cloning and expression in Escherichi
A;Reference number: A34355; MUID:90036724; PMID:2681162
A;Accession: A34355
A;Status: preliminary

RESULT 5

T14850
S-layer protein precursor - Bacillus stearothermophilus
C;Species: Bacillus stearothermophilus
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T14850
R;Jarosch, M.; Egelseer, E.M.; Mattanovich, D.; Sleytr, U.B.; Sara, M.
submitted to the EMBL Data Library, April 1999
A;Description: Nucleotide sequence of the coding region of sbsC, the S-layer gene from B
A;Reference number: 218239
A;Accession: T14850
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1099 <JAR>
A;Cross-references: EMBL:AF055578; NID:g4581955; PID:g3025826; PIDN:AAC12757.1
A;Experimental source: ATCC 12980
A;Genetics:
A;Gene: sbsC

Query Match   6.1%;   Score 255.5; DB 2; Length 1099;
Best Local Similarity   22.0%;   Pred. No. 0.00023;
Matches 214;   Conservative 125;   Mismatches 330;   Indels 303;   Gaps 49;

Qy   58 VDENAQPSFKDAKNIWS-----------------SKYIAAVEKAGVVKGDG-----------------KEN 94
                                        : : :|||: ||:|: |:
Db   56 VTETGQ--FPDIKDVYAAYNKAKQAYANAVAVVNKAGGAKKDAYLADLQAIYETVFKAN   113

Qy   95 FYPEG---------KIDRASFASML-----------------VSAYNLK-------------------   117
                          |||:||||| |:| |||
Db 114 --PKSGEARVATYIDAYNYATKLDKMRQELKAAVDAKDLKKAEELYHKISYELKTRTVIL   171

Qy 118 DKVNGE------LVTTF----------------------------EDLLDHWGEEKANIL---IN   145
              |:| :| |:|| |:|||||
Db 172 DRVVGQSTRELLRSTFKADAQALRDRLLYDITVAMKAREAQDAVKAGNLDKAKAALDQVN   231

Qy 146 LGISVGTGG-KWEPNKSVSRABAAQFIALTDKKYGKKDNAQAYVTDVKVSEPTKLTLTGT   204
           |||| | |: ||:||| : :||:| |:: : |:|||:||:|
Db 232 QYYSKVTDAFKAELQKAAQDAKAAYEAALTPK--------------VESVSAIDSTSFKVTFT   280

Qy 205 -GLDKLSA-----DDVTLEGDKA---VAIEASTDGTSAVVTLGGKVAPNKDLTVKVKN--- 253
          :|||: : |:::|||:| :|:|:|||||::|||||:||||| |:||||:
Db 281 KPVDKATAIPKNFSITLKGTETKLYPKSVEVEBSGLTATVTLYDTILVDGKTYTVVTSGLK 340

Qy 254 ------QSFVTKPVEYEVKKLAVEKLTPDDDRAGQAIAFKLNDEKGNADVEYLNLANHDV 306
                :||| | ||:|| : :::|| |: ||:|:||::| :||: |::
Db 341 DTAGKEFETSTNEFTY-NKPVPASIIFNFNKLPEDSAVDL--------TKYVT------V 385

Qy 307 KFVANN---------LDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLTVSNT 359
          |:|::   ::|:   :|||: |:| :::|| : |: ::|:||||:|:
Db 386 KDAAGNVIKSGFELEFTSSEKLTQGKF-NTTGKKS---------VIVNAIVKG--TNVTT 434

Qy 360 GIITVKNLDTPASAIKNVVFALDADNDGVVNY---GSKLSGKDFALNSQNLVVGEKAS-- 414
          || :|| :|||::| :| :| :|| :|| : |:||:|:| ||| ||:|||::
Db 435 GNVILAVEDEKAAEVSEL-KLTKDNKEVVTLYANGNAFDKDGNQJSSGTLTLTAKPKDQ 492

Qy 415 -LNKLVATIAGEDKVVDPGSISIKSSNHGIISVV-NNYITAEAAGEATLITKVGDVTKDV 472
          : |||| :|||:|:| ||||||||:|:||||:| |:||:|||||::|| |||:|||:
Db 493 YGNELTGKVAGTD------YTFESLNPEVLVVAPDGSVTPIVPGTALVKVKYGEVTKTI 545

Qy 473 KFKVTTDSRKLVSVKANP----------DKLQVVQNKTLPVTFVTT--DQYGDPFGANTAAIK 523
          ||||||||||||||||| ||||||||||||||||| ||||||||||||||
Db 546 P----VTVKANFVLETIAVDSTGVSVAKGQXATFKVTLKDQYGNKFTGNVNVTS   595

Qy 524 EVLPKTGVUAEGGLDVVTTDSGSIGTKTIGVTGNDVGEBGTVHFQNGNGATLGSLYVNVTE   583
          |:|||||:|||||||::|||:|:||||||||||||::||||:||||:||:||||:|:|:||
Db 596 D---KTETA------TVSVSNSGIGQSEYTVTVNGVAEGSTTITIKSGTKEVKVPVNVVA   646

Qy 584 GNVAFKNFELVSKVGGVQGYGQSPDTKJDLNVSTTVEYGLSKYTSDRVVSDPENLEGYVESK 643
          |||||||||||||||:|||||||||:|||||:|||||||||||:|||:|||:|:|||||||:
Db 647 GG-PVANYQI--KV-----LDDGKID----------KSATESPANNDVQLKVYAVDAN   686

Qy 644 NLAVADAKIVGNKVVTGKTP----------------GKVDIHLTKNGA-TAGKATVEIVQ   687
          ||:|:||:|||||:||||| :|:|: | :||| ||||:||:|:|
Db 687 GNIVGD---ITNDVTIISEATDTNGVIVNASKSTANGDTVVVITDNGSKKVGKETLITVKL 743

Qy 688 ETAIKSVNFKPVQT--ENFVEKKINI--------------GTVLELEKSNIDDIVKGIN----- 731
          |:|:||:|| |:| |||:|||:| ||||:|||:||:|||||:|
Db 744 GTVTLGTVDVEVIDTTLKATVVTKKADLIELDAADNGDALAKLLANLD--IKDQNGNPMV 801

Qy 732 --LTKETQHKVRVVK-----------------------------SGABQGKYLDRNGDAVPNAG-- 765
            |:||| ||:|:| ||: :|:|||::|:|| :||
Db 802 DSAATPNTNEKLQALKSVLSGIVSSDTSVIGSVSNVDNLKDDASIGGLAVKKAGTVTLTL 861

Qy 766 ----DVKLGDVTVSQTSDSALPN-----FKADLYDTLT---TKY-TDK--GTLVFKVL 808
              | |||||:|::||||| | ||||||||:: ||| || || |:|:|
Db 862 VFNEDSKIAPIAITVKAPAATQDGVTVTGLDLIVPGVTGVGKTKFTATDKIKSGHKLYYAV 921

Qy 809 KDKDVITSEIGS 820
          |||||||:|||
Db 922 DDSAVPAPAVGT 933

RESULT 6
S29043
cellulase (EC 3.2.1.4) - Bacillus sp.
N;Alternate names: endo-1,4-beta-glucanase
C;Species: Bacillus sp.
C;Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 01-Dec-2000
C;Accession: S29043; PC4404
R;Ozaki, K.; Shikata, S.; Kawai, S.; Ito, S.; Okamoto, K.
J. Gen. Microbiol. 136, 1327-1334, 1990
A;Title: Molecular cloning and nucleotide sequence of a gene for alkaline cellulase from
A;Reference number: S29043; MUID:91037937; PMID:2230718
A;Accession: S29043
A;Molecule type: DNA
A;Residues: 1-941 <OZA>
A;Cross-references: EMBL:M27420; NID:g142664; PIDN:AAA22304.1; PID:g142665
R;Shirai, T.; Yanane, T.; Hidaka, T.; Kuyama, K.; Suzuki, A.; Ashida, T.; Ozaki, K.; Ito
J. Biochem. 122, 683-685, 1997
A;Title: Crystallization and preliminary X-ray analysis of a truncated family A alkaline
A;Reference number: PC4404; MUID:98060488; PMID:9339567
A;Accession: PC4404
A;Molecule type: protein
A;Residues: 228-584 <SHI>
A;Experimental source: strain KSM-635
C;Function:
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A;Pathway: cellulose degradation
C;Superfamily: Bacillus sp. KSM-635 alkaline cellulase; S-layer repeat homology; Thermot
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;41-95/Domain: S-layer repeat homology <SLR1>
F;101-153/Domain: S-layer repeat homology <SLR2>
F;164-219/Domain: S-layer repeat homology <SLR3>
F;766-908/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA>

Query Match   6.0%;   Score 252.5; DB 2; Length 941;
Best Local Similarity   21.3%;   Pred. No. 0.00025;
Matches 193;   Conservative 109;   Mismatches 291;   Indels 311;   Gaps 42;

Qy   5 FPDVPAGHWAEGSINYLVDKGAITGKPDGTVGPTESIDRASAAVIFTKILNLPVDENAQP 64
         |||||| ||::||:|||||||:||| ||||||| ||:|||:| |:|:| |||::|:|:||
Db   41 FSDVKKTSWSFPVIKDLYEQEVITGITSATTFSPTDSVTRAQFTVMLTRGLGLEASSKDYP 100

Qy   65 SFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNL-------- 116
         |||:||:||||||:|| |||:|||:|:||| |||:|||::| |:|||||||
Db 101 -FKDRKN-WAYKEIQAAYEAGIVTGKTNGEFAPNENITREQMAAMAVRAYEYLENELSLP 158

Qy 117 --KDKVNGELVTTF-EDLLDHWGEEKANILINLGISVGTGKWEPNKSVSRAEAAQFIA 172
           |||:||||||| |:|||||||::||:|||: ||:|||||:||:|:||||:|:|::
Db 159 EEQREYNDSSSISTFAQDAV----QKAYVELM--EGNTDGYFQPKRNSTREQSAKVIS 211

Qy 173 LTDKKYGKKDNAQAYVTDVKVSEPTK------LTLTGTGLDKLSADDVTLEGDKAVAIE- 225
          |||||||||||||:||:|:||:||| :|:|||||||:| ||||:|||:|||:
Db 212 TLLWKVASHD--YLYHTEA-VKSPSEAGALQLVELNG----------QLTLAGEDGTPVQL 259

Qy 226 --ASTDGT--SAVVTLGGKVAPNKD---------LTVKVKNQSFVTK-----FVYEVKKLA 268
           |:|||| ||||||||||:|||| |||:|:|||||:| |||:|||||
Db

```
Db  260  RGMSTHGLQWFGEIVNENAFVALSNDWGSNMIRLAMYIGENGYATNPEVKDLVVYEGIELA  319
Qy  269  VEKLTF--------DDDRAG-----QAIAFKLNDEKGNADVEYL--NLANHDVK         307

Db  320  FEHDMYIVIVDWHVHAPGDPRADVYSGAYDFFEIADHYKDHPKN---HYIIWELANEP-    374
Qy  308  FVANNLDGSP-ANIFPGGEATSTGKLAVGI-KQGDYKVEV--QVTKRGGLTVSNTG       360

Db  375  --SPNNNGPGLTNDEKGWEAVKEYAEPIVEMLREKGDNMILVGNPNWSQRPDLSADNP-    431
Qy  361  IITVKNL------------DTPAS-----AIKNVVFALD---------              382

Db  432  -IDAENIMYSVHFYTGSHGASHIGYPEGTPSSERSNVMANVRYALDNGVAVFATEWGTSQ   490
Qy  383  ADNDG-------VVNVGSK--LSGKDFALNSQNLVVG--EKASLNKLVATIAGEDKVV     429

Db  491  ANGDGGPYFDEADVWLNFLNKHNISWANWSLTNKNEISGAFTPFELGRTDAT------DL   544
Qy  430  DPGSISIKSSNHGIISVVNNYITAEAAGEATLTIKVGDVTKDVK--------FKVT       477

Db  545  DPGANQWAPEE--LSLSGEYVRARIKGIEYTPI--DRTKFTKLVWDFNDGTTQGFQVN     599
Qy  478  TDS--RKLVSVKANPDKLQV----VQN---------------                     498

Db  600  GDSPNKESITLSNNNDALQIEGLNVSNDISEGNYWDNVRLSADGWSENVDILGATELITID  659
Qy  499  ------------------KTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVV           532

Db  660  VIVEEPTTVSIAAIPQGPAAGWANPTRAIKVTEDDFESFGDGYKALVTITSEDSPSLETI   719
Qy  533  AEGGLD--------VVTTDSGIGTKTIGVTGNDVGEGTVHFQNGNGATLGSLY-----     578

Db  720  ATSPEDNTMSNILFVGTEDADVISLDNITVSGTEIBIEVIHDEKGT-ATLPSTFEDGTR    778
Qy  579  -------------VNVTEGN------------VAFKNFELVSKVGQ                 599

Db  779  QGWDWHTESGVKTALTIBEANGSNALSWEYAYPEVKPSDGWATAPRLDFWKDELVRGTSD   838
Qy  600  Y------------------GQSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEG       637

Db  839  YISFDFYIDAVRASEGAISINAVFQPPANGYWQEVPTTPEIDLITELDSATVTSD-ELYH   896
Qy  638  YEVE                                                           641

Db  897  YEVK                                                           900
```

```
Db   54  LNLPVDENAQ--PSFKDAKNI-WSSKYIA-AVEKAGVVKGDGKENFYPEGKIDRASFA    107
Qy   80  LEIPMDERAAQTLAAKFKDAADIADWAQRQVAVAVEK-GYVKGYPDGTFLPNADLNGLAFC 138

Db  108  SMLV------SAYNLKDKVNGELV------                                 125
Qy  139  SLILQQLGYDGDFVFDERAYKLQEFGGLTAEQABAFNNKNGINRDSMVGIAFSALQAVYK  198

Db  126  ---TTFEDLLDHWGEEKANILINLG-------ISVGTGGKWEP------              158
Qy  199  ATGKTVIEVLVEN-GNVSKELAIELGVLLKAIKEVKALDAVKVQVGK---EPVLPEEVE    253

Db  159  --NKSVSRAEAAQFIALTDKKYGKKD------NAQAY-------VTD                190
Qy  254  VVYEDDTEKLAVEWPTVDTSEVGEQEIEGTIKGASGLAYREPKATLKVIVTPEELQVVD    313

Db  191  VKVSEPTKLTLTGTGLDKLSADD---VTLEGDKAVAIEASTDGTSAVVTLGGKVAPNKDL   247
Qy  314  VKAPNLKEIVIEFNGEVASKADEKSSYSVEDNTIELVTVSEDKTTVTLTVAGAMTAEEEI   373

Db  248  TVKVKNQSFVTKFVYEVKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLNLA---NH    304
Qy  374  EVTIKT---ATGLKEEVTKTVVPA----------DYENPEAEBSIALIGPNSF          412

Db  305  DVKPVANNLDGSPANIPEGG---------EATSTTGKLAV-GIKQGDYKVEVQVTK-      350
Qy  413  EIKFSEPVQSSSDAEVLVNDGTYYVSEEKLSQDYRTLTVELGVSSLNEGTYKVKVKGYRD  472

Db  351  -RGGLTVSNT-GIITVKNLDTPASAIKN----------VVFALDADNDGVVNYGSKLSGKDFA 401
Qy  473  YAGNIMRTKTFDLEYVKDTTPTAKVKEATQNKVVIEFNEPATRDG-------YSGDEAA   525

Db  402  LNSQ---------NLVGEKASLNKLVATIAGEDK-------VVDPGSISIKSSNH       441
Qy  526  LTRDYFQTYSSWKPTKVVASDN---NKVYTLYFSEDQNDGGYPVVLLPVGNVTIT-----  578

Db  442  GIISVVNNYITAEAAG--------EATLTIKVGDVTKDVKFKVTTDSRKLVSVKANPDKLQ 494
Qy  579  -ILKEVDDDAVVDAWGNKLESDLKLTATVAADNEAPTVK-SVTAEAE--------DKIV   627

Db  495  VVQNKTLPVTFVTTDQYGDPFGANTAAIKE-VLPKTGVVAEGGLDVVTIDSGSI------  547
Qy  628  VV----------------FSEDVNENQAKDKDNYVIKKDGKEIDTAISSITYDSNETKVTIV 673

Db  548  ---------------GTKTIGVTGNDVGEGTVHFQ---------------             567
Qy  674  LDEKLSGGKYTIDIKGIKDTSVSENEMKAVTIEFEVTDKIAPTIEEVTFVDNYIYRYSE   733

Db  568  ----NGNGATLGSLYVNVTEGN-----VAFKNFELV-SKVGVGQYGQSPDTKLDLNVSTTVEYQ 619
Qy  734  AMSTKGNGSVLRKDNYKLVDDNDKKVEIKKLELFGSDKNKVRITVDSDVDLN-----VDYZ 789

Db  620  LSKYTSDRVYSDPEN-LEGYEVESKQLAVADAKIVGNKVVVTGKPGKVDIHLTKNGATA   678
Qy  790  L---TIANVEDEAGNAISAFDVKAKKLSEEQAPEV-SEIRIISKI--EIEIVINK--IL   840

Db  679  GKATVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLELEKSNLDD---IVKGINLTKE  735
Qy  841  DKATVE----------------KTDFEVERGSNKVALTRISSITYDDGKTIVKGV------ 879

Db  736  TQHKVRVVKSGAEQGKLYLDRNGDAVFNAGDVK----------LGDVTVSQTSDSALPNFKADL 789
Qy  880  LPDAVRPANSG------DITGYTLYIVGEIKSDTGKEMATGAVSKPVDD---KFAPSF    928

Db  790  YDFLTTKYTDKGTLVFKVLKDKDV--ITSEIGSQAVHVNVLN                    829
Qy  929  VSVANGVYGDASKKGFLTFDEDIKPLNNSAGLGATDLVIKN                     970
```

RESULT 7

Query Match   5.8%; Score 242.5; DB 2; Length 1036;
Best Local Similarity 19.4%; Pred. No. 0.00086;
Matches 206; Conservative 152; Mismatches 351; Indels 353; Gaps 52;

```
Qy    8  VPA-----GHWAEGSINYLVDKGAITGKPDGYGPT--ESIDRASAAVIFTKI-------    53
Db   22  VPAPADSFSYEKEAEILYRL--GLYKGTSETEVPNLEGKLDRQTGVVMLLRLFSQEDDA    79
```

RESULT 8

```
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T31105
R;Ward, C.K.; Lumbley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.
J. Bacteriol. 180, 6013-6022, 1998
A;Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.
A;Reference number: Z20984; MUID:99030326; PMID:9811662
A;Accession: T31105
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4919 <WAR>
A;Cross-references: EMBL:AF057696; NID:g3929021; PID:g3929023; PIDN:AAC79761.1
C;Genetics:
A;Gene: lspA2

Query Match           5.7%;    Score 240.5; DB 2;  Length 4919;
Best Local Similarity 20.4%;  Pred. No. 0.0099;
Matches   206; Conservative 147; Mismatches 353;  Indels 305; Gaps 51;

Qy    13 WAEGSINY-----LVDKGAITGKP--DGTYGPTESIDRASAAVIFTKILNLPVDENAQPS   65
Db   294 FAAGNLTYDVNTRDVNRNTNPKPRPITDNTRKDNIAISGESAGSMYGRNIKFIVTDKG---  350

Qy    66 FKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPE--GKIDRASFASMLVSAYNLKDKVNG  122
Db   351 ------------AGVNHQVIFAEDDINILTDDGNSRLNKVYADYVRVVGKDIELANNG   397

Qy   123 ELVTTFEDLLDHWGEEKAN-----ILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKY  178
Db   398 QIHADQQLIINATGHVKLNDGSSVISNNNLGISALNLTLENATVS-ANNLSPVTNNTKL  456

Qy   179 GKKDNAQAYV-----------------------TDVKVSEPTKLITGTGLDK--------  208
Db   457 NNLSKVSARAADLQSGNLNLDKASVLAHKLTLNISNDVSLNNQSKLSANNLKIKKVRDLN  516

Qy   209 -----LSADDVTLEGDKAVAIEASTDGTSAVVTLG--GKVAPNKDLIVKVKNQSF-VTKF  260
Db   517 LNNSELSANNLTLNTSNNITLKNKSKFTAGNMTLNVTNNVTLNNDSELAANNLTLNVTKN  576

Qy   261 V--YEVKKLAVEKLTFD--DD-------RAGQAIAFKLNDEKGNADVEYLNLANHDVK  307
Db   577 VTLNDASKLSANKLDLNVTDNVTLNSKSTLSAGE-LTFK--------KVKNVTL-NNDSE  626

Qy   308 FVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLTVSNTGIITVKNL  367
Db   627 LAANNLS--------LNASHNVTLNNKSKLSAQKADIKAV-------NLTLNDTTELTAKNL  673

Qy   368 DTPASAIKNVVFALDADNDGVVN----YGSKLSGKDFAL----NSQNLVV------  409
Db   674 DINSTTI-------TNNGTIAGIFANTTEKLNNKEKALILAEQNLNFTVNGSHYENK  724

Qy   410 GEKASLNKLVATIAGEDKVVDPGSISIKSSNHGIISVVNNY---------ITAEAA  456
Db   725 GDIVSKDKATVTFSKNSDFTSNGSKLVNAQNQLKVN-VNNFTISQGDDITLIGNVTLNAS  783

Qy   457 GEAT---------LTIKVGD--------VTKDV----KFKVTTDSRKLVSVK-ANPD  491
Db   784 GTFTNSGNLTTVKTLDVGDIQNFTNKGNLTVGEDLHIKSKTKTNDG-KLISIKNLNISS  842

Qy   492 KLQVVQNKTL---PVTFVTTDQYGDPFGANTAIKEVLPKTGV----VAEG------  535
Db   843 EADFINNGTLLGIEIALKIATKGNF-------TNKEKAILAASNLLDISVAEGKKTFNNGT  895

Qy   536 ---GLDVVTTDSG--------SIGTKTIGVTGNDVGEGTV-------------HFQNG  569
Db   896 IESGKNLNIITNGAFLNVDNATIRSFGVLNITSTGNVSNGTLISNERLNITSAANFTNE  955

Qy   570 NGATL---GSLYVNVTEGNVAFKNFELIVSKVGQY--GQSPDIKLDLNVSTTVEY-----  618
Db   956 SNGTVMSNGLLNIIAKQGNITNKN--LIASRQQLNLTAVADNITNDSNISNKIAVLHSLG  1013

Qy   619 QLSKYTSDRVVSDPENLEGYEVESK-NLAVADAKIVGNKVVVTGKTPGKVDIHLTKNGAT  677
Db  1014 NISLNSKDQVVNLGEIYAGNNISVKAHQLKNDVKLMGD---IITKT-----------K  1057

Qy   678 AGKATVEIVQETIAIKSVNFKPVQTENFVBKKINI-GTVLELEKSNLDDIVKGINLTKET  736
Db  1058 EGQASYKLVQ--ASNGGHFGNDGSSGYSEGDANIKGKFADD------------NKLT  1101

Qy   737 QHKVRVVKSGAEQGKLYLDRNGDAVFNAGD-------VKLGDVTV---SQTSDSALPN-  784
Db  1102 VQRI--------GKIYAGR--DLTFNKSNAGGKSEIINRGTINVKNKLSYDSDVSFENN  1150

Qy   785 ---FKADLYDTLTKYTDKGTLVFKVLKDKDVITSEIGSQAVHVNVLNNPN  832
Db  1151 MQSQKVDLY----------TKIFEAKSDIELJFKTNGTHPVYLNFKSNNN  1190
```

RESULT 9

```
B59102
hypothetical protein pX01-90 - Bacillus anthracis virulence plasmid pX01
C;Species: Bacillus anthracis
C;Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 11-May-2000
C;Accession: B59102
R;Okinaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koehle
J. Bacteriol. 181, 6509-6515, 1999
A;Title: Sequence and organization of pX01, the large Bacillus anthracis plasmid harbor
A;Reference number: A59091; MUID:99445483; PMID:10515943
A;Accession: B59102
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-652 <OKI>
A;Cross-references: GB:AF065404; NID:g4894216; PIDN:AAD32394.1; PID:g4894306
A;Experimental source: strain Sterne
A;Note: similar to hypothetical protein; PFB065w (980 aa); Plasmodium falciparum (AE00
C;Genetics:
A;Gene: pX01-90
A;Genome: plasmid

Query Match           5.7%;    Score 239.5; DB 2;  Length 652;
Best Local Similarity 23.8%;  Pred. No. 0.00061;
Matches   114; Conservative 80; Mismatches 188;  Indels 97; Gaps 19;

Qy     4 SFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRAS-AAVIF-TKILNLPVDEN   61
Db    44 AFLDVPSNHWANVAINNLLKKEIIVGYGNNKFGLGDSVTREQVAAVIYRIKFPEKEVEAE  103

Qy    62 AQPSFKDAKNIWSSKY---IAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKD  118
Db   104 AQNPYRDV-NKSSTMFLDEILTJTKMGIFSGDEKGNFRPKSPISRAEMTQVIKQNAFNI-  160

Qy   119 KVNGELVTTFEDLL-DHWGEEKANILINLGISVGTG-GKWEPNKSVSRAEAAQFI  171
Db   161 PVVGK--HNFRDVAKDYWANDAISALQSNQLVSGTGNGLFEPYKSVIREEYAQFIFNVLN  218

Qy   172 ---ALTDKKYGKKDNAQAYVTDVKVSEPTKLILTGTGLDKLSADDVTLE------GDKAV  222
Db   219 FEDLVNKEKQLDELNKRRDVINQKIDE-------------FDKLNSQRKDLERMLELNQKLS  268

Qy   223 AIEASTDGTSAVVTLGGKVAPNKDLTVKV-KNQSFVTKFRVYEVKKLAVEKLTFDDDRAGQ  281
Db   269 QLKQQ--SPQLQDLKNLKBSQSRLLELNKKDSNRLELNSEIKKL------NDRKAE  317

Qy   282 AIAF-------KLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIFEGGEATS  328
Db   318 LLSLIMELIKQQSEFDKKIKNEKDDLNKKREDLIN--------------RIAESKELAK  362

Qy   329 TTGKLAVGIKQGDYKVEVQVTKRGGLTVSNTGII--------TVKNLDTPASAIKN---  376
Db   363 KKAELNTKLVE-LFKVQEALNKKSGQYLYYINKLDNELRELADKYKNSDNKISRLKNHIG  421

Qy   377 -------VVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLNKLVATTIAGEDK  427
Db   422 BYNKQLEKIENELECNKKIDNTKKQLABFDKSNKKQQELESELVQLNKKIDELGKRHK  480
```

RESULT 10

```
A56143
surface-array protein homolog sapA2 - Campylobacter fetus (strain 82-40 LP3)
C;Species: Campylobacter fetus
C;Date: 03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999
C;Accession: A56143
R;Dworkin, J.; Tummuru, M.K.R.; Blaser, M.J.
J. Bacteriol. 177, 1734-1741, 1995
A;Title: A lipopolysaccharide-binding domain of the Campylobacter fetus S-layer protein
A;Reference number: A56143; MUID:95204338; PMID:7896695
A;Accession: A56143
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1109 <DWO>
A;Cross-references: GB:S76860; NID:g913763; PIDN:AAB33871.1; PID:g913764
A;Note: this gene appeared to be silent in strain 82-40 LP3
C;Genetics:
A;Gene: sapA2

Query Match       5.6%;     Score 236;    DB 2;    Length 1109;
Best Local Similarity  21.6%;    Pred. No. 0.0019;
Matches 208;   Conservative 130;   Mismatches 361;   Indels 266;   Gaps  49;

Qy   25 GAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQSFKDAKNIWSSKYIAAVEKA  84
Db  218 GVGTAAESTLNPGDKIDGGAG---NDVLKVDLKNN----FKGLKD---DGYIKNIEKL 266

Qy   85 GVVKGDGKENFYPEGKIDRASPASML----VSAYNLKD----KVNGELVTTF------E 129
Db  267 SLTNSSVSNRTFDAKGIDGLQTVALSGEKGISVTNLANIVDVEVNGFKGTNFNVDSIYAD 326

Qy  130 DLDDHWGEEKANILIN---LGISVG-TGGKWEP------------------------NKS 161
Db  327 KVLD-GSADVQNLKVNGVGAKGASVAITADKIETLNNTTGSQSFVSADVASISVKGNAN 385

Qy  162 VSRAEAAQFIALTDKKYGKKDNA-----QAYVTDVKVSEPTKLITGTGLDKLSADDVTLE 217
Db  386 LSLATGAKTTTLDASSFGGALDADLSTSASVTSIK-------GGNGNDKITIIKDVAVN 436

Qy  218 -----GDKAVAIEASTDGT--------SAVVTLGGKVAPNKDLIVKVKNQSFVTKFVVE 263
Db  437 VAIDGGAGNDELVIKGSTADTLQPTLTNIBKVTIDGN--TKDLTLSLKKAQSVTELSFK 493

Qy  264 -----------VKKLAVEKLTFDDDRAGQAIAFKLNDEKG-NADVEYLNLANH 304
Db  494 NIAKTVTESNGNVETVNILANNATDKAVTINDESLKTINFSDVDDKGASVAAKGKIVADK 553

Qy  305 DVKFVANNLDGSPAN-IFEGGEATS---TTGKLAVGIKQGDYKVEVQVTKRGGLTVSNT 359
Db  554 ATELTINSNKVTLASDAVVQAANATKIDINAAKDTVGLTLGG-----VAKLTDLITVNNK 607

Qy  360 GIITVKNLD-TPASAIRNVVFALDADNGVVNYGS--------KLSGKDFALNSQNLV 408
Db  608 GAFALTGANATDLDSVKN----LSVNTEGAFSIATATSLKNLNNLSLNGVSADLNSVNVG 663

Qy  409 VGEKASLNKLVATIAGEDKV---VDPGSISIKSSNHG---------------IISVVNN 449
Db  664 TATLASLEANI-NVSGEFKLGTTTAKGDVDFNIENVGALITLGAITSSTGNASVIISSATG 722

Qy  450 YIT--AEAAGEATLTIKVGDVTKDVKR----------------KVTTDSRKLVSVKANP 490
Db  723 NVTLGAVSATOGNLTLNAGNTLGNITIGALKGDIVSVDLGGVLGTINSDANNKVSITSN- 781

Qy  491 DKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVAE--GG--LDVVTTDSGS 546
Db  782 ------EVTYV----GSEISKNVVEITAAAGGTDLNAQVIGGAAADDALTIIGK 825

Qy  547 IGTKTIGTGNDVGEGTVHFQNGNGATLGSLYNVTE---GNVAFKNFELVSKVGGYQGS 603
Db  826 GDTQTTTASG-DLSGGTLTLTLTEATKLSSLDISGVKGITGNVAI------ELGKAVOG 877

Qy  604 PDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEGYEVEKSNLAVADAKIVG-----NKV 657
Db  878 --NKTDVSVQGSDAAEOITYTS----------------AASLTDIKISGDLGAGANTI 917
```

```
RESULT 11
AC1533
surface protein (LPXTG motif) [imported] - Listeria innocua (strain Clip11262)
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; M
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland
A;Title: Comparative genomics of Listeria species.
A;Reference number: A81077; MUID:21537279; PMID:11679669
A;Accession: AC1533
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1386 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC96035.1; PIDN:g16413254; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin0803

Query Match       5.6%;     Score 233.5;   DB 2;    Length 1386;
Best Local Similarity  21.9%;    Pred. No. 0.0034;
Matches 197;   Conservative 121;   Mismatches 339;   Indels 243;   Gaps  47;

Qy   27 ITGKPD--GTYGPTESIDRASAAVIFTKILNLPVDEN-----AQPSFKDAKNIWSSK- 76
Db  556 VTGEVDVNTPGDYEIMYTIDGVSTKIIVT----VKEDKSSIEAKDSILYIGDTWNSKD 609

Qy   77 -YIAAVEKAGVVKGDGKENFYPEGKIDRASPASMLVSAYNLKD-KVNGELVTT--FEDL 131
Db  610 NFISATDKDG-------------------------NPVDFKDIKVEGTVNTIKPGTNKV 643

Qy  132 LDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYGKGKDNAQAYVTDV 191
Db  644 TYLYGNQSKEVTITV----------------KADQSTLEAKDSIIYTGDKWNAKDN----- 683

Qy  192 KVSEPTKLTLTGTGLDKLSAD--DVTLEGDKAVAIEASTPGTSAVVTLGGKVAPNKDLTV 249
Db  684 ---------FITATDKDGNPVDFKDIEVEG-----TVDTTKPGTNKITYIYGNL--SKEVTV 729

Qy  250 KVK-NQSFVTKFPVEVKKLAV--------EKLTFDDDRAGQAIAFKLNDEKGNAD--- 295
Db  730 TVKANQA------TLEAKDSALYVGDTWNSKDNFISATDKDGTPVDFKDIKVEGTVDTTKA 784

Qy  296 ----VEYL--NLANH-----DVKFVANNLDGSPANIFEGGE------ATSTTGKLAVGI 337
Db  785 GTNKVTLYGNQSKEVTVNVKADQTLEVKDSVIYTGDKWAEDNFISATDKTGN-SVGF 843

Qy  338 KQGDYKVE--VQVTKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDAD-----NDGVV 389
Db  844 K--DIKVEGTVDTTKAG---------TNKVTYTYANQSKEVTIVKADQATLEAKDSII 891
```

```
Qy   390  NYGSKLSGKDFALNSQNLVVGEKASLNKLVATIAGEDKVVDPGSISIKSSNHGI------  443
          ||  |||||:| ||: :    :|   :       ||:
Db   892  YTGDKWKAEDNFISATD-KAGKTIDFKNI--KVEGTIDTTKAGNYDITYSYSGVTRSTEL  948

Qy   444  -----ISVNNYITAEAAGEATLTIKVGD--VTKDVKFKVTTDSRKLVSVKANPDKLQVV  496
                 :|||  |:   |||:|| :: |||||:|||:|:|   :   :: :
Db   949  SKTITVTVKKNQVNLEAKDS--TLYEGDKWIAKDNFVSATDKDGNTVDFKAIEVKGTV-  1004

Qy   497  QNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVA-----------EGGLDVVTTD-  543
          :|  :|       ||: |:
Db   1005 -NTTKAGTYKITYSYAGI-SKITVTVTVLANQTKIVAKDLIIYEGDNWKEQDNFVSATDK  1061

Qy   544  -SGSIGTKTIGVTGN-DV--GEGTVHFQGNGATLGSLYVNVTEGNVAFKNFEL-----  593
           ||| ||: |||:|  |:    ||: ::| |:| |  :| :|  ||
Db   1062 FGQAIDFNSVKVTGSVDIQTPGKYRITYSIEGASTTITVTVLADQSNLVAKNSTIYVGDK  1121

Qy   594  -----VSKVGQYGQSPDTKLDLNVSTVE----------YQLSKYTSDRVYSDPENLE  636
               |:|||  ||||:| |||:||||:            :|:|||   |   : :
Db   1122 WQSKDNFVSATDKYGKPIDLSL-LTVTGVDTTTPGEYBITYSVNGLTTTIVTVKENQA  1180

Qy   637  GYEVESKNLAVADA-KIVGNKVVTGKT----------PGKVDIHLTKNGATAGKATVE  684
          | ||| :| | |: |:| :  ||||           |  |||: ||||: |:
Db   1181 NIVABDSTIHAKESWKAADNFVSATDKTGKNIGLSSVVTGKVDINTPGNYEIT--YTID  1238

Qy   685  IVQETIAIK-SVNFKPVQTENFVEKKINIGTVLELEKSNL----------DDIVKG-  729
           |||:|| | ||| ||| |||||:||:||| ||:|  ::||
Db   1239 GVSTTITVTVLVNHSQIEAH---DAKIKVGTSWKPEDSFILAKDKFGETADFSDVTVEGT  1295

Qy   730  INLTKETQHKVRVVKSGAEQG-KIYLDRNGDAVFNAGDV-----KLGDVTVSQTSDSALP  783
          |||:|:|:||   |: :|:::  :|:| |:   ::|  :     ||  : :|:|:|:
Db   1296 VDT-TKPGKYQITYTIDGVSVTITVIVEDNSKGSSNTTTPPNHSIGEEETSNSNDKQLP  1355
```

RESULT 12
D82671
surface protein XF1529 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: D82671
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: D82671
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2059 <SIM>
A;Cross-references: GB:AE003982; GB:AE003849; NID:g9106554; PIDN:AAF84338.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigo
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Palmieri, D.A
, F.G.; Nunes, L.R.; Rosa, V.E.; de Oliveira, M.C.; de Oliveira, R.C.; de Sa, R.G.; Santelli, R.V.; Sawasak
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silvei
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1529

```
Query Match             5.5%;  Score 231.5;  DB 2;  Length 2059;
Best Local Similarity  21.4%;  Pred. No. 0.0075;
Matches 193;  Conservative 129;  Mismatches 338;  Indels 241;  Gaps 43;

Qy   32  DGTYGPTESIDRASAAV--IFTKILNLPVDENAQPSFKDAKNIWSSKYIAAVEKAGVVKG  89
         ||   |    :  :     :: : |:|  ::  :||:||||   :|:: ||::
```

```
Db   834  DGPAVTASGIDAGSKVISHVAAGVVSETSTDAVNGSQLNAVQVQASOPVTFTGNEGAVKR  893
          ||| |::||||||||||| ||::|  |: :|| |  :: ||::|:     ||||:| |||
Qy   90   DGKENFYPEGKIDRASFASMLVSAYNLK---DKVNGELVTTPEDLLDHWGEEKANILINL  146

Db   894  SLGQSVVISGESSTAG----TYSGGNLKSVVDEAAGRIHLQLAD-----SPKFGNVVINN  944
          :|||||:||| :| |        | |||::| ||:|| ||:|:|
Qy   147  G--ISVGTCGKWEPN------KSVSRAEAAQFIALTDKKYGKDNAQAYVTDVKVSEPT  197

Db   945  GGKISGVTAGTEETDAVNFSQLKSISTA-VDQGWTLTASGANGSKVASGGTVDLKNTD-G  1002
           || ||| |:||||||||||:|::|:|   ||||||:|||:|||||| ||| ||::| |
Qy   198  KLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVTLGGKVAPNKDL---TVKVKNQ  254

Db   1003 NLTISKSG----DSNDVVFNLSKDFKVDGMTSGTTVVNNDGVKVGSDVALGTTGLTIANG  1058
          ||:|||        || || |||||: |:|| ||:||:|  |  :|:: |
Qy   255  SFVT-----KFVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLNLANHDV  306

Db   1059 PAVTASGIDAGSKVISHVAAGAVSETSTDAVNGSQLNAVQVQ----------ASQPV  1105
          ||::||||||||||| ||:||:|  |: :||:|  |::||||:
Qy   307  KFVAN---------NLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVT---  349

Db   1106 TFTGNEGAVKRSLGQSVVISGESSTAGTYSGGNLKSVVDEAA--------GRIHLQLADSP  1158
           |||| ||||| ::|:|||| ||: |  |   || :|:
Qy   350  KRGGLTVSNTGII--TVKNLDTPA---SAIKNVVFALD---------ADNDGVVNYGSK  394

Db   1159 KFGNVVINNGGKISGVTAGTEETDAVNFSQLKSISTAVDQGWTLTASGANGSKVASGGT-  1217
           |||:||:| |||| ||| |:|||||||||||:|| |||:|||||::|||||||:| |:
Qy   395  LSGKDFALNSQNLVVGEKASLNKLVATIAGEDKVVDPGSISIKSSNHGI-----  443

Db   1218 --VDLKNTDGNLTISKSGDSNDVVFNLSKDFKVDGMTSGTTVVNNDGVKVGSDVALGTT  1274
            |||:|||||| ||||||||||||||||||:| |:||| ||:|:|  | :|:: |
Qy   444  -ISVVNN-YITA---EAAGEATLTIKVGDVTKDVKFKVTTDSRKLVSVKANPDKLQVVQ-  497

Db   1275 GLTIANGPAVTASGIDAGSKVISHVAAGAVS-----ETSTDA------VNGSQLNAVQV  1322
          ||:|:||||:||||||||||||| ||:||:       :::|
Qy   498  NKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVVTTDSGSIGTKTIGVTGN  557

Db   1323 QASQPVTFT---------GNEGAVKRSL---------VAFKNFELV--------  594
          ||:||:|           ||| | ::
Qy   558  DVGE---GTVHFQGNGATLGSLYVN-----VTEGN------VAFKNFELV--------  594

Db   1365 VVDEAAGTIHLQLADSPKFGNVVINNGGKISGVTAGTEETDAVNFSQLKSISTAVDQGWT  1424
           |:|||| || ||  |   |:|:|| |:  || |: || ::| ||:|
Qy   595  -------SKVGQYGQSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENL-----------  635

Db   1425 LTASGANGSKVAS-GGTVDLK--NTDGNLTISKSGSDNDVVFNLSEDLKEKSITVGNTQ  1480
          ||:|| |||||:  |||||:    ||||||||||:|:||||||||:||: ::||:|:| :
Qy   636  ------EGYEVESKNL----------------AVADAKIVGNKVVTGKTPGKVDI-----  669

Db   1481 LDKDGVKVSSNVLLDSNELVITSHSSTSSVKTLANGESVVNRTVVNGDGVNIDDVVVVND  1540
          |   ::|   :  || :|| :|  :: :::|:   |::|::|  |:|| |:||| ::|:|
Qy   670  ---------HLTKNGATAGKATV-EIVQETIAIKSVNFKPVQTENFVEKKINIGTVLELE  719

Db   1541 LGLSIVGGASLTLSGINAGSHKITNVTAGTEDTDAVNFSQLKS---VSEAVDKGWTLTAS  1597
          |||:| ||:::|:|||| |:|||| |:|||:||||||||:|||     |:||||||||:|
Qy   720  KSNLDDIVKG--INLTKETQHKVRVVKSGAEQGKLYLDRNGDAVFN-AGDVKLGDVTVSQ  776

Db   1598 GANGSKVVSGGTVDL-KNTDGNLAISKSG--------DSN-DVVFNLSKDFKVDEVTAGN  1647
          ||||||| ||||||: ||||||| |||||        ||  ||||||||||| |||  :
Qy   777  T  777

Db   1648 T  1648
```

RESULT 13
B99789
hemagglutinin/hemolysin-related protein [imported] - Escherichia coli (strain O157:H7,.
C;Species: Escherichia coli
C;Accession: B99789
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
R.Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: B99789
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <HAY>
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs1282

```
Query Match          5.5%;    Score 231;  DB 2;   Length 1268;
Best Local Similarity   21.9%;   Pred. No. 0.004;
Matches  200;  Conservative  139;  Mismatches  358;  Indels  218;  Gaps   51;

Qy      6 PDVPAGHWAEGSINYLVDKGAIT-GKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQP 64
            ||:|   |:|:||    |:||:|:|                   |:||:|
Db    159 PDIQDDKLA---GYSVNGGTITIGK-----------LDNASP----TEILSRNVVVNGKV 200

Qy     65 SFKDAKNIWSSKYI-AAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGE 123
            |::||||||:|||||             ::|||||             :||::|
Db    201 SADELNVVAGNNYVNAAGQVTGSVSATGSRNGY---SVDVAKLGGMYA-----NKIS-- 249

Qy    124 LVTTFEDLLDHWGEEKANILINLGI------SVGTGGKWEPNKSVSRAEAAQFIALTDKK 177
            |||:|                   |     ||:|             ::|:|
Db    250 LVST------------EKGVGVRNLGVIAGGVNGVSIDSKGNLLNSNAQIQSASTINLTTN- 298

Qy    178 YGKKDNAQAYVTDV-KVSEPT-KLTLTGTGLDKLSADDVTLEGDKAVAIEAST-DGTSAV 234
            |:|::|:            |   |:|:|::|     :| ||||:||::||
Db    299 -GTLDNTTGTVTSVGTISLNTNKNTIVNT-----RAGNISTMGD--IYVNSGTIDNTN-- 348

Qy    235 VTLCGKVAPNKDLTVKVKNQSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAFK---LNDEK 291
            |:||||             ::| :|:|  |||||:|                |:|::|
Db    349 ----GKLAAAGMLAVDTNNATLINS---GKGSSVGIE-----------AG-LVALKTGTLNNSN 393

Qy    292 GNADVEYLNLANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKR 351
            |:|:            |:| :| |:|  |: ||              |:|:|: :|
Db    394 GQIRGGYVGLES-----AALNNNNGD------IQTTGDIAI-ISNGN-----VDNN 432

Qy    352 GGLTVSNTGIITV-------------KNLDTPASAIKNVV--FALDADNDGVVNYGSKLSGK 398
            |||:||||||:|             |||||:            |:||:||
Db    433 KGLIRSSTGHIVIGAAGSVNNGSTKTADTGSSDSLGIIADTGVEIGANNINNNGGQIA-- 490

Qy    399 DFALNSQNLVVGEKASLNKLVATIAGEDKVVDPGSISIKSSNHGII---------SVV 447
            |||||:||  |:|:|               ||:|||:|:
Db    491 ---SNGNVSLSSYSTIDDYAGKILSNSKVIIKGS-SLRNDTGGISGKQGIEVAVGGSLT 545

Qy    448 NNYITAERAAGEATLTIKVGDVTKDVKF----KVTTDSRKLVSVKANPDKLQVVQNKTLPV 503
            |||:|||:|||             |:|      |:|:||:||||:|||||||:|||
Db    546 NN-IGVISSEEGDISLLANSVDNHGGFMMGQNITMESMSGV---NNNTALIVASKKLKI 600

Qy    504 TF--VTTDQYGDPFGANTAAIKEVLPKT--GVVABGGLDVVTTDSGSIGTKTIGVTGNDV 559
            ||  |:|||:|::||     |::|||| |||:|||||:|||:||||:|:|:||:|||||
Db    601 NARGSIENRDGNNFG-NAYGLYFGMPQQTGGMVGKEGIELSGQNIYNNSRLIAEDGPLT 659

Qy    560 GEGTVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQYGQSPDTKLDLNVSTTVEYQ 619
            ||||:||||||:||| |:|::||:||||||||||||||||||||||||:||||:||:||:
Db    660 LQAQNTFDNTRALVTSGADASIQVGGTYYNNYATTWSAG-----------NLDID-ATTLQNS 710

Qy    620 LSKYTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVDIHLT-KNGATA 678
            |:| :|||::|::|:|     |:|||   |:||     |:||:||||||:|:| :||||
Db    711 SSGTMID-----NNATGFIASDKNLSLE----VVNSLTNYGWISGKGDVDTVNNGNLY 760

Qy    679 GKATV--EIVQETIAIKSV-NFKPV-----------QTENFVEKKINIGTVLEL 718
            |:||:  |||::|:||:|: ||||             |::||:|:|:||:|:|
Db    761 NRNTIAAEKGLDIAALNGIENWKDISAGGDLTMNTNRHVTNNSNMVGQNIVINAVNDI 820

Qy    719 EKSNLDDIVKGINLTKETQHKVRVVKSGAEQGKLYLDRNGDAVENAGDVKLGDVTVSQTS 778
            |||||:||:||:|||||||||           |::|:|:||   |:|:|||:|::|::
Db    821 --NRNGNIVSDADL------NVTTKGNLYNYLYMVGYGDIALSANSVANNATIEATG 870

Qy    779 DSALP------NFKADLY---DTLTTK----YTDKG-------TLVFKVLKD--KDVIT 815
            |:|||      ||:||:|   |:|:||    |:|:|       ||:|||||:  ||||:
Db    871 DLIIDSKGNVGNNRGNLHALNGVLSVKGNNLNNDNGEIRGYGDVTLALTGNYDSYKGSLT 930
```

```
Query Match          5.5%;    Score 231;  DB 2;   Length 1270;
Best Local Similarity   21.9%;   Pred. No. 0.004;
Matches  200;  Conservative  139;  Mismatches  358;  Indels  218;  Gaps   51;

Qy      6 PDVPAGHWAEGSINYLVDKGAIT-GKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQP 64
            ||:|   |:|:||    |:||:|:|                   |:||:|
Db    161 PDIQDDKLA---GYSVNGGTITIGK-----------LDNASP----TEILSRNVVVNGKV 202

Qy     65 SFKDAKNIWSSKYI-AAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGE 123
            |::||||||:|||||             ::|||||             :||::|
Db    203 SADELNVVAGNNYVNAAGQVTGSVSATGSRNGY---SVDVAKLGGMYA-----NKIS-- 251

Qy    124 LVTTFEDLLDHWGEEKANILINLGI------SVGTGGKWEPNKSVSRAEAAQFIALTDKK 177
            |||:|                   |     ||:|             ::|:|
Db    252 LVST------------EKGVGVRNLGVIAGGVNGVSIDSKGNLLNSNAQIQSASTINLTTN- 300

Qy    178 YGKKDNAQAYVTDV-KVSEPT-KLTLTGTGLDKLSADDVTLEGDKAVAIEAST-DGTSAV 234
            |:|::|:            |   |:|:|::|     :| ||||:||::||
Db    301 -GTLDNTTGTVTSVGTISLNTNKNTIVNT-----RAGNISTMGD--IYVNSGTIDNTN-- 350

Qy    235 VTLCGKVAPNKDLTVKVKNQSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAFK---LNDEK 291
            |:||||             ::| :|:|  |||||:|                |:|::|
Db    351 ----GKLAAAGMLAVDTNNATLINS---GKGSSVGIE-----------AG-LVALKTGTLNNSN 395

Qy    292 GNADVEYLNLANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKR 351
            |:|:            |:| :| |:|  |: ||              |:|:|: :|
Db    396 GQIRGGYVGLES-----AALNNNNGD------IQTTGDIAI-ISNGN-----VDNN 434

Qy    352 GGLTVSNTGIITV-------------KNLDTPASAIKNVV--FALDADNDGVVNYGSKLSGK 398
            |||:||||||:|             |||||:            |:||:||
Db    435 KGLIRSSTGHIVIGAAGSVNNGSTKTADTGSSDSLGIIADTGVEIGANNINNNGGQIA-- 492

Qy    399 DFALNSQNLVVGEKASLNKLVATIAGEDKVVDPGSISIKSSNHGII---------SVV 447
            |||||:||  |:|:|               ||:|||:|:
Db    493 ---SNGNVSLSSYSTIDDYAGKILSNSKVIIKGS-SLRNDTGGISGKQGIEVAVGGSLT 547

Qy    448 NNYITAERAAGEATLTIKVGDVTKDVKF----KVTTDSRKLVSVKANPDKLQVVQNKTLPV 503
            |||:|||:|||             |:|      |:|:||:||||:|||||||:|||
Db    548 NN-IGVISSEEGDISLLANSVDNHGGFMMGQNITMESMSGV---NNNTALIVASKKLKI 602

Qy    504 TF--VTTDQYGDPFGANTAAIKEVLPKT--GVVABGGLDVVTTDSGSIGTKTIGVTGNDV 559
            ||  |:|||:|::||     |::|||| |||:|||||:|||:||||:|:|:||:|||||
Db    603 NARGSIENRDGNNFG-NAYGLYFGMPQQTGGMVGKEGIELSGQNIYNNSRLIAEDGPLT 661

Qy    560 GEGTVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQYGQSPDTKLDLNVSTTVEYQ 619
            ||||:||||||:||| |:|::||:||||||||||||||||||||||||:||||:||:||:
Db    662 LQAQNTFDNTRALVTSGADASIQVGGTYYNNYATTWSAG-----------NLDID-ATTLQNS 712
```

Db 1173 TYQVSGL---KNGDTAGSILTGGLNRAAGENVGV--YG--INQGDLALNSGNYDLSYQGN 1225

Qy   412 KASLNKLVATIAGEDKV-------VDP----------------------------GSI 434
          ::|: | ||| |:
Db 1226 NLTITKALLNVIADAKTKVYGDADPSLTYQVSGLKNGDTAGAVLNGGGLVRVSGENVGNY 1285

Qy   435 SIKSSNHGIISVVNNYITABAAGEATLTIKVGDVTKDVKFKVTTDSRKLVSVKANPDKLQ 494
          :|:   :| ||:| |:|
Db 1286 AIQQGGLGLVS--GNYDLAYQGNNLTITKALLNVIADAKTKVYGD--------ADPS--- 1332

Qy   495 VVQNKTLPVTF-VTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVVTTDSGSIGTKTIG 553
          :| : ||:|| | || | | |: |
Db 1333 ------------LTYQVSGLKNGD--------SAGSILTGGLNRAAGEN-----VG 1363

Qy   554 VTGNDVGEGTVHFQN----------GNGATLGSLYNVTEGNVAFKNFELVSKVGQYGQSPDTK 607
          |||: | | |: |:|:| |: | |:|| |||:|| |
Db 1364 VYGINQGDLALNSGNYDLSYQGGNNLTITKALLNVIAD--------AKTKVYG----- 1407

Qy   608 LDLNVSTTVEYQLSKY----TSDRVYS------DPENLEGYEVESKNLAVA----DAKI 652
          ||| | : |:|:   :|: |: :|| | |||
Db 1408 -DADPSLT--YQVSGLKNGDTAGAVLNGGGLVRVSGENVGNYAIQQGGLGLVSGNYDLAY 1464

Qy   653 VGNKVVVT----------GKTPGKVDIHLT----------KNGATAGKATVEIVQETIAIKSV 695
          ||| :| || ||:| | |:||: | | |:|:|:||:||
Db 1465 QGGNLTITKALLNVIADAKTKVYGDADPSLTYQVSGLKNGDTAG--------AVLNGG 1514

Qy   696 NFKPVQTENFVEKKINIGTVLELEKSNLDDIVKGINLTKETQHKVRVKSGABQGKLYLD 755
          |: |||:|| :|| | :| | |::| |:||:|:|:|||||||||:|||||:||
Db 1515 SLSRVAGENVGVYGINQGD-LALNSGNYDLSYQGNNLT-ITKALLNVIADA--KTKVYGD 1570

Qy   756 -------------RNGD---AVFNAGDV-----KLGDVTVSQ----------- 776
                        |:| |||:   :||||||:|
Db 1571 ADPSLTYQVSGLKNGDTAGAVLNGGGLVRVSGENVGNYAIQQGGLGLVSGNYDLAYQGNN 1630

Qy   777 -TSDSALPNFKADLYDTLTTKYTDKG-TLVFKV--LKDKD----VITSEIGSQA---VHV 825
           |:| || | |: |: |||:| ||||: |:| |:: : |:|
Db 1631 LTITKALLNVIA---DAKTKVYGDADPSLTYQVSGLKNGDSAGSILTGGLNRAAGENVGV 1687

Qy   826 NVLNNPNL 833
          |: | |:|
Db 1688 YGINQGDL 1695

Search completed: April 7, 2004, 17:33:00
Job time : 29 secs

Qy   620 LSKYTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVDIHLT-KNGATA 678
          ::| :|| |:|:| |||:| |:|||||:|| |||| ||:||: |:|||
Db   713 SSGTMID------NNATGFIASDKNLSLE-------VVNSLTNYGWISGKGDVDVTVNNGNLY 762

Qy   679 GKATV--EIVQETIAIKSV-NFKPV---------------QTENFVEKKINICTVLEL 718
          ||| | : | |||:|| |::| |||||||||| ::|||||
Db   763 NRNTIAAEKGLDIALNGIENWKDISAGGDLTMNTNRHVTNNSNSNMVGQNIVINAVNDI 822

Qy   719 EKSNLDDIVKGINLTKETQHKVRVVKSGABQGKLYILDRNGDAVFNAGDVKLGDVTVSQTS 778
          |||:|||:|| :: ||:|| ||| |:||||:|:|||:|:|
Db   823 --NRGNIVSDADL--------NVTTKCNLYNYLYMVGYGDIALSANSVANNNATIEATG 872

Qy   779 DSALP----NFKADLY---DLTTK---YTDKG------TLVFKVLKD-KDVIT 815
          ||| | ||| | ||:|| | ||||:| |||:
Db   873 DLIIDSKGNVGNNRGNLHALNGVLSVKGNNLNNDNGEIRGYGDVTLALTGNYDSYKGSLT 932

Qy   816 SEIGSQAVHVNVLNN 830
          |::|| || | |::|
Db   933 SETGDVTLTANIVDN 947

RESULT 15
F83068

Query Match   5.5%;   Score 230.5; DB 2;   Length 2154;
Best Local Similarity 22.3%; Pred. No. 0.0089;
Matches 216; Conservative 112; Mismatches 327; Indels 313; Gaps 49;

Qy   43 RASAAVIFTKILNLPVDENAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKID 102
          :|: |: | : | | |: :| :| |||:||||:|:|||
Db   864 QATASTTATQASN--VGQYAITGNANGSEYFSQRY------QLVRQDGRLTVTP----- 909

Qy   103 RASFASMLVSAYNLKDKVNGELVTTFE------DLLDHWGEEKANIILNLGI 148
          ||: | | || |||| :|| ||::||| |||:| ||
Db   910 ---AQLIVSA-DAKTKVYGDADPTLITYQVSGLKNSDTAAGVLSGNLGRVAGENVGNYGI 964

Qy   149 SVGTGGKWEPNKSVS-----RAEAAQFIALTDKK---YGKKDNAQAY-VTDVKVSEPTK 198
          |:|:||| | || | ||: |: |: || ||:|| || |:|:|| |:|
Db   965 LQGGLGLNTANYTLSYVGNDLRITPAQLNVIADAKTKVYGDLDPALTYQVSGLKRGDTAG 1024

Qy   199 LTITGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVTLGGKVAPNKDLTVKVK-NQSFV 257
          |||| || :|:|||| | ||| ||| || | | | | |:|:| |:| |
Db 1025 AVLNGGSLSRVAGENVGVYG------------INQGGLGLVSSNYTLNYQGNNLTI 1068

Qy   258 TKFVYEVKKLAVEKLTFDDDRA--GQAIAFKLNDEKG---------------NADVEYLN- 300
          ||| | || ||:| ||:||| |||||||:|||:: |||| |
Db 1069 TKALLNVIADAKTKVYGDADPALTYQVSGLKNGDTAGAVLNGGSLSRVAGENVGVYGINQ 1128

Qy   301 ------LANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGL 354
                ||:| | |: ||| | |:|||| | || |:|:|:|| |
Db 1129 GGLGLLSANYDLSYQGNNL-----------TITKALLNV---IADAKTKVYGDADPSL 1172

Qy   355 TVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLLSGKDFALNSQNLVV---GE 411
          || |:||| ||| |: |:| || | ||| ||:| | || |||:|| |: : ||

```
                    Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:        April  7, 2004, 17:24:49 ; Search time 51 Seconds
                                        (without alignments)
                                        5153.464 Million cell updates/sec

Title:         US-09-844-281-1
Perfect score: 4202
Sequence:      1 AGKSFPDVPAGHWAEGSINY.........ITSEIGSQAVHVNVLNNPNL 833

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      1017041 seqs, 315518202 residues              1017041

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     SPTREMBL_25:*
               1: sp_archea:*
               2: sp_bacteria:*
               3: sp_fungi:*
               4: sp_human:*
               5: sp_invertebrate:*
               6: sp_mammal:*
               7: sp_mhc:*
               8: sp_organelle:*
               9: sp_phage:*
              10: sp_plant:*
              11: sp_rodent:*
              12: sp_virus:*
              13: sp_verebrate:*
              14: sp_unclassified:*
              15: sp_rvirus:*
              16: sp_bacteriap:*
              17: sp_archeap:*

                        SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

Result No.	Score	% Match	Query Length	DB	ID	Description
1	765.5	18.2	821	2	Q9RED0	Q9red0 bacillus th
2	730.5	17.4	816	2	Q9ZES5	Q9zes5 bacillus th
3	706	16.8	823	2	Q45759	Q45759 bacillus th
4	564.5	13.4	529	16	Q81UH4	Q81uh4 bacillus an
5	556.5	13.2	414	16	Q81S57	Q81s57 bacillus an
6	528.5	12.6	483	16	Q813K8	Q813k8 bacillus ce
7	528	12.6	531	2	Q8KYE8	Q8kye8 bacillus an
8	509	12.1	344	16	Q81U01	Q81u01 bacillus an
9	506	12.0	410	16	Q81QA4	Q81qa4 bacillus an
10	504	12.0	410	16	Q81AN7	Q81an7 bacillus ce
11	475	11.3	530	16	Q81HB4	Q81hb4 bacillus ce
12	395	9.4	931	2	Q9L6S5	Q9l6s5 bacillus ps
13	367	8.7	920	2	Q45664	Q45664 bacillus st
14	322	7.7	1268	2	Q9RER7	Q9rer7 bacillus sp
15	307.5	7.3	1047	2	Q9RB35	Q9rb35 cytophaga s
16	281	6.7	3133	2	Q8KRR1	Q8krr1 fusobacteri

17	278	6.6	510	16	Q81Y62	Q81y62 bacillus an
18	263.5	6.3	459	16	Q81S56	Q81s56 bacillus an
19	263	6.3	766	2	Q9EXR6	Q9exr6 bacillus sp
20	257	6.1	3165	16	Q8RDQ9	Q8rdq9 fusobacteri
21	256.5	6.1	461	16	Q81PJ9	Q81pj9 bacillus an
22	256	6.1	2937	16	Q8D990	Q8d990 vibrio vuln
23	255.5	5.9	1099	2	O68840	O68840 bacillus st
24	250	5.9	2143	16	Q8RIP5	Q8rip5 fusobacteri
25	249.5	5.9	2353	2	P71401	P71401 haemophilus
26	245	5.8	1292	2	Q841Y6	Q841y6 campylobact
27	243.5	5.8	1881	16	Q8RGK2	Q8rgk2 fusobacteri
28	243.5	5.8	2806	16	Q8RI19	Q8ri19 fusobacteri
29	243	5.8	1557	2	Q9RNI2	Q9rni2 haemophilus
30	243	5.8	3692	2	Q8KRR3	Q8krr3 fusobacteri
31	242.5	5.8	1036	2	O86999	O86999 clostridium
32	241.5	5.7	2402	2	Q9AER7	Q9aer7 staphylococ
33	240.5	5.7	4919	16	Q92HL0	Q92hl0 haemophilus
34	240	5.7	2444	16	Q81GX1	Q81gx1 bacillus ce
35	239.5	5.7	652	2	Q9X360	Q9x360 bacillus an
36	238	5.7	1299	16	Q9F3X6	Q9f3x6 pasteurella
37	236	5.6	1109	2	Q53505	Q53505 campylobact
38	235	5.6	2712	16	Q9F3X5	Q9f3x5 pasteurella
39	234	5.6	4254	16	Q7UU38	Q7uu38 rhodopirell
40	233.5	5.6	1386	16	Q92DL0	Q92dl0 listeria in
41	233.5	5.6	1794	16	Q8RHH1	Q8rhh1 fusobacteri
42	231.5	5.5	360	16	Q81N79	Q81n79 bacillus an
43	231.5	5.5	2059	16	Q9PD50	Q9pd50 xylella fas
44	231.5	5.5	8173	16	Q7UDU8	Q7udu8 rhodopirell
45	231	5.5	1270	16	Q8XAN9	Q8xan9 escherichia

```
                        ALIGNMENTS

RESULT 1
Q9RED0
ID   Q9RED0;                PRELIMINARY;       PRT;       821 AA.
AC   Q9RED0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Surface-layer protein precursor.
GN   SLPA.
OS   Bacillus thuringiensis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1428;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 4045;
RX   MEDLINE=21220765; PubMed=11320137;
RA   Mesnage S.; Haustant M.; Fouet A.;
RT   "A general strategy for identification of S-layer genes in the
RT   Bacillus cereus group: molecular characterization of such a gene in
RT   Bacillus thuringiensis subsp. galleriae NRRL 4045.";
RL   Microbiology 147:1343-1351(2001).
DR   EMBL; AJ249446; CAB63252.; -.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF00395; SLH; 3.
KW   Signal.
FT   SIGNAL        1       29
SQ   SEQUENCE     821 AA;   87279 MW;   8D68995C81221488 CRC64;

Query Match        18.2%;  Score 765.5;  DB 2;  Length 821;
Best Local Similarity  32.7%;  Pred. No. 2.5e-23;
Matches 284;  Conservative 114;  Mismatches 316;  Indels 155;  Gaps 42;

Qy      1 AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE 60
            |:|||||| :|:||||:|:||:||||:|||||||||:||||||||:||||||:|||
Db     30 AGKTFPDVQPGSWSABYIDYLVAKKAIEGKPDG TFAPTEAIDRASAAKIMAITLGLEVKD 89

Qy     61 NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENPYPEGKIDRASFASMLVSAYNLKDKV 120
            | :|||||:|::||:||||||| :|||| |||:|||| ::||| ||:| |: :|
Db     90 GAKPSFKDAQDSWAAXYIAAVEQAGVIQGDGTGNFNPSNQINRASMASMIVKAYKLEDKV 149
```

```
DR   InterPro; IPR003343; Big_2.
DR   InterPro; IPR008964; Invasin_intimin.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF02368; Big_2; 1.
DR   Pfam; PF00395; SLH; 3.
DR   SMART; SM00835; BID_2; 1.
KW   S-layer.
SQ   SEQUENCE   816 AA;  87293 MW;  100289290580215l CRC64;
```

```
Query Match             17.4%;  Score 730.5;  DB 2;  Length 816;
Best Local Similarity   30.6%;  Pred. No. 6.4e-22;
Matches 269;  Conservative 119;  Mismatches 321;  Indels 169;  Gaps 36;

Qy    1 AGKSFPDYPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE 60
          ||||||:|||:||  ::||||||:||:|||  ||:|  |:|||:|||:  ||:|||||:|
Db   30 AGKSFPDVPADHWGIDSINYLVEKGAVTGNDKGMFEPGKELTRAEAATMMAQILNLPIDK 89

Qy   61 NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV 120
          :|:|||  :  |::  :|||||||||:|| | |||:|||||||:| |::||:||:||.|||
Db   90 DAKPSFADSQGQWYTPFIAAVEKAGVIKGTG-NGFEPNGKIDRVSMASLLVEAYKLDTKV 148

Qy  121 NGELVTTFEDLLD-HWGEEKANILINLGISVG-TGGKWEPNKSVSRAEAAQFIALTDKKY 178
          ||  :|:|:||| |||.||||||:| |||||| |.||||| ::|::|||||||| |||:
Db  149 NGTPATKFKDLETLNWGKEKANILVELGISVGTTADKWEPKKTVTKAEAAQFIAKTDKQF 208

Qy  179 G----KKDNAQAYVT--DVKVSEPTKLTLTGTGLDKLSADDVTL---EGDKAVA--IE 225
              : |:|| :|:   |::| :      ::   :||:.:||:| :   |.||||      
Db  209 GTEVAKVBSAKAVTTQKVEVKFSK--------AVEKLTKSDVKLANKANNDKVLVKDVK 259

Qy  226 ASTDGTSAVVTLGGKVAPNKDLTVKVK--NQSFVTKFPVYBVKKLAVEKLTFDDDRAGQAI 283
           :::  :|| ::  :.||  ::|:| |  |:|:| | |: ||:  :|    ::.  .:|:
Db  260 LSEDKKSATVELYSNLAAKQTYTVDVNKVGKVEVTVGSLEAKTIEMADQTVVADEP-TAL 318

Qy  284 AFKLNDEKGNADVEYLNLANHDVKFVANNLDGSPANI-FEGGEATSTTGKLAVGIKQGDY 342
           :|:.|:.|:   :|||::|:      |: :.||  :  |  |:   :.| | :||  .:
Db  319 KYTVKDENGTEVV-----------------SPAGIEFVTPAABKINAKGEITTLAKGTS 359

Qy  343 KVEVQVTKRGGLTVSNTGIITVKNLDTPASAIKN-VVFALDAD------N 390
          .: | : |  |:|.:|.: :|: :  :. |::.  |::| .:|      :
Db  360 TTVKAVYKDGKVVAESKEVKVSAEGTAVASISNWTVAAEKADFTSKDFKQNDKYYEGDN 419

Qy  391 YGSKLSGKDFALNSQNLVGEKASLNKLVATIAGEDKVVDPGSISIKSSNHGIISV---- 446
            |::| :|   |:|:||.|| ||||.|:::.| :||   | |||:|||:|.|||:
Db  420 VSVQVELKDQFNNVVNNVKAEYESLNTEVAVV---DKAT-GKVTVLSAGKAPVKVTVKD 474

Qy  447 -----VNNYITAEA-AGEATLTIKVGDVTKDVKFKVTTDSRKLVSVKANPDKLQVVQNK 499
               |:|::|.||  :|||:|:|| |||||||||||||||||||||  :||  :.:
Db  475 SKGKELVSKTVEIEAFAQKAMKEIKLEKTNVALSTKDVTDFK-----VKA--------- 519

Qy  500 TLPVTFVTTDQYGDPFGANTAIKEVLPKTGVVAEGGLDVVTTDSGSIGTKTIGVTGNDV 559
           :|||:||:||||||.:  |:|.| ::  | |:|:   :  : .:|||: ||||||.||
Db  520 --PV---LDQYGKEFAAPVEV---KVLDKDGKELKEQKLVAKYENKEL--VLNAHGQEA 568

Qy  560 GEGTVHFQNGNGATLGSLYNVTEGNVAFKNFELVSKVGQVGQSPD--TKLDLNVSTTVE 617
          |: ||:   :.|:  |:: :: .|  |:| | |::| |:|:. .   ||| |||||||:
Db  569 GKYTVELTAKSGKK------------EVKSKLALELKAPGVFSKFPVR--GLE 607

Qy  618 YQLSKYTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVTGK------TPGKVDIH 670
            |:|.||:: |::| :|:|| |||:||||||||||||||||||||      |.|||||
Db  608 NELDKYVTEENKKNEMVVSVLPVDANGLVLREKEAATLKVTTDKDGKVVDATSGQVAVN 667

Qy  671 ----------LTKNGATAGKA-TVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLELEKSNL 723
                    |||||||||| ||||||||||||||||||||||||||||||||||  :::
Db  668 DAAGTITVGNEAKAGEYKVTVVADGKLIITHSFKVVDTAPAAKK-----LAVDFTSTSL 722

Qy  724 DDIVKGINLTKETQHKVRVRVKSGAEQGKLYLDRNGDAVFNAG----DVKL--GDVTVS 775
          :::: | |:|::  ::| : .: |:: |: : | ::  .:  .    |:|:   |:||:
Db  723 NEVAQGSEL------KTALLN------iiSVDGVPATTAGATVTDVKFVSADTNVVSE 768

Qy  776 QTSDSALPNFKADLYDTLTTKYTDKGTLVFKVLKDKDV 813
          :::  |  ||||||||| :::|.|||:|:|: |: |::
Db  769 ETAKFGTKGSTSIFVKELTVK---KGEQTQKVELDKPV 803
```

```
Qy  121 NGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAAQPIALTDKKYGK 180
          ||  :|:|:||| ||||||||||:| ||:||.|:.|||:.||.||| || :|| ||::.|:
Db  150 SGELETNFSDLKDHWGEKDANILVALGINGTGNGWEPDKSVTRAEEAAKFIAKTDMQFGQ 209

Qy  181 KDNAQAYVTDVDVKVSEPTKLTLT-GTGL-DKLSADDVTLEGDKAV--AIEASTDGTSAVV 235
          |:.|:| ::: :||:  :  :::| |.:  :|:::...:||.|:::  :::|:.|.:|||
Db  210 K--AEAKVESIKEINAKEIEVKFGTEVKDVTAANFAVVEGSKELDIEKVELSKDKKSATI 267

Qy  236 TLGGKVAPNKDLTVKVKNQSFVTKFVYEVKKL-------AVEKLTFDDDRAGQAIAFKLN 288
          || :::  | |::||:|.:..||  :|:||:        :|::| .::|:||:|:|||:|
Db  268 TLKNAL-------VNKQAYVA-HVKDVKSVDGKDIPKALEVIFFNDEVAPTVSVTSTP 317

Qy  289 DEKGNADVEYLNLANHDVKFVANN---LDGSPANIFEGGEATSTTGKLAVGIKQGDYKVE 345
          |:.|:| :|  |:|  :| |:|::   |.|:| .: |.|: :|:| |||.  | |  :|
Db  318 D--GNVKVVFSEKLSKDAVTVVINGKEFTATP---BENTVLTKADVASVVKNGE-AFN 370

Qy  346 VQVTKRGGLTVSNT----GIITVK-NLDTPASAIKNVVFALDADNGGVVNVGSKLSGKD 399
          |:||.|  ||..|     | :||: ::|.|.::||:: |: .:|.|.   ||:: |:
Db  371 VIVTGAKDL-VGNTMEMYEGKATYKVEKDVTAPEVKDI-KVKELVDGVATLEVTFSEE- 426

Qy  400 FALNSQNLVGEKASLNKLVATIAGEDKVVDPGSISI-KSSNHGIISVNNYITAEAAGE 458
           :|:::.:||.||:|:||:|::.||||::  |::|: ||||:|||||:|:||||||||
Db  427 --LSAQGKVVVKKG----------DKVVEGATIALDQDTTKALVTVPGALTDKETA- 470

Qy  459 ATLTIK-VG--DVTKDVKFKVTTDSRKLVSVKANPDKLQVVQNKTLPVTFVTTDQYGDPF 515
          |:|::| ||  :|||:|||||:|| ::| |:||:|:.|::|:: |::|:|:|:::.||:|
Db  471 ANLKVEFVGYKDAANNVGNKVTKEVKVTKDVVA-PNLVKVVADENKAATF-TFD------ 522

Qy  516 GANTAAIKEVLPKTGVVAEGGLDVVTTDSGSIGTKTIGVTGNDVGEG--TVHFQ-NGN-- 570
          |::.| :|:|:.|:|::.:.||:::::||:.:.||:|::|:|:.:.:  |::|: :|:
Db  523 ------KEV------TAQEGKLRVINLDTSKDVTKEVAVASVEDNKKAITLIFQBKGNYK 570

Qy  571 -GATLGSLYNVTEGN--VAF-KNFELVSKVGQYGQSPDTKLD--LNVSTTVEYQLSKYT 624
           .|:|.:|:.|.:|:  .|| |:|:||:|||:.::    :|   |||||:||.|:|.||
Db  571 VAATKG--FVKDTAGNESAAFTKEVKVVEK------KEEGKKDEVAPKATKVERVADSKT 622

Qy  625 SDRVYSDPENLEGYEVES----KNLAVADAKIVGNKVVVTGKTPGKVDIHL--------- 671
          |:.|.::|:.::|:. :|    ||||||||||||||||||||||||||:||
Db  623 KFTVTFDKEVKGGQGADSASNVNNYTLAGAKLPEGTLIVVNADGKSVTIELPETFTPEKS 682

Qy  672 --------TKNGATAGKATVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLELEK 720
                  |:|:|:.|: ::::.::|:|::.|:|:|.:|:||::||:.|::|::|:||
Db  683 ETVKFTVANVANKDGVKMG--TTNLLVNVVDTKAPEFKSAKITKVDAKEITL-TFSEAVN 739

Qy  721 SNLDDIVKGINLTKETQHKVRVRVKSGAEQGKLYLDRNGDAVFNA---GDVKL--GDVTVS 775
          |::::::| |:|:::  ::| : .: |:::|: : | ::  .: |   |:|:   |:||:
Db  740 VDATDFVIDLNGVVLTATK--ADETAEASK------DVVLKVTAPADVNLATGTVTVK 789

Qy  776 ---------QTSDSALPNFKADLYDTLT 794
                   :::  |  ||||||||| :
Db  790 AKEVEGKVVLNTADTSINKNKLIVAFKPVT 818
```

RESULT 2
Q9Z8S5

ID   Q9Z8S5   PRELIMINARY;   PRT;   816 AA.
AC   Q9Z8S5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CTC protein.
GN   CTC.
OS   Bacillus thuringiensis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1428;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CTC;
RA   Sun M., Yu Z.;
RT   "S-layer protein gene of Bacillus thuringiensis CTC.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ012290; CAA09981.1; -.

Db   596 KEVKVVEKEAE--KETDKVAPVKAVAYDKATNKVT---VNFDKEVKGGQVAESAANVNN 650
            ::| :|:| : | | | .: | | |: || : |
Qy   645 LAVADAKIVGNKVVVTGKTPGKVDIHLT----------------KNGATAGKATVEIV 686

Db   651 YTLAGAKLPEGTLIVLDGTKAIIELPSTFTFEKSETVKFTVANVANKDGVKMG--TANLL 708
            :| ||| |||:|:||||| ||:||||:|| |: |:|:||| | :||
Qy   687 QETIAIKSVNFKPVQTENFVEKKINIGTVLELEKSNLDDI----VKGINLTKETQHKVR 741

Db   709 LEVKDTKAPEFKSAKITKVDAKEI---TLTFSEAVNVDTDFTVDLNGVELT------VA 759
            || :|||||| |||||||||:||:| :|| |:|:|||:|:||||||| |:
Qy   742 VVKSGAEQGKLYLDRNGDAVFNA---GDVKLGDVTVSQTSDSALPNFKADLYDTLTTKYT 798

Db   760 KADATAEAAK-------DVVLKVTAPTDVNLATGTVTVKAKELENNAK--YEFKTTDKA 809
            | |:| :| | |:| | | ::|:||||||||:||||||| |||||||
Qy   799 DKGTLVFK 806

Db   810 ENALVAFK 817

RESULT 4
Q81UH4
ID   Q81UH4   PRELIMINARY;   PRT;   529 AA.
AC   Q81UH4;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   N-acetylmuramoyl-L-alanine amidase, family 3.
GN   BA0898.
OS   Bacillus anthracis (strain Ames).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=198094;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22608414; PubMed=12721629;
RA   Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA   Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA   Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA   Kolonay J.F., Beanan M.J., Dodson R.J., Durkin A.S., Haft D.H.,
RA   Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA   Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA   Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA   Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA   Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA   Fraser C.M.;
RT   "The genome sequence of Bacillus anthracis Ames and comparison to
RT   closely related bacteria.";
RL   Nature 423:81-86(2003).
DR   EMBL; AE017027; AAP24895.1; -.
DR   TIGR; BA0898; -.
DR   GO; GO:0008745; F:N-acetylmuramoyl-L-alanine amidase activity; IEA.
DR   GO; GO:0009253; P:peptidoglycan catabolism; IEA.
DR   InterPro; IPR002508; Amidase_3.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF01520; Amidase_3; 1.
DR   Pfam; PF00395; SLH; 3.
DR   SMART; SM00646; Ami_3; 1.
KW   Complete proteome.
SQ   SEQUENCE   529 AA;   57941 MW;   CB0AC553BD215AE9 CRC64;

Query Match   13.4%;
Best Local Similarity   32.6%;   Pred. No. 2e-15;   DB 16;   Length 529;
Matches 172;   Conservative   81;   Mismatches 192;   Indels   83;   Gaps   18;

Qy   1 AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE 60
           |||||||| .|.|:. |::|||.||| ||||||| ..: |:|.||| ||| :.||||:
Db   25 AAKKFSDVPT--WAQQSVDYLVGKKALDGKPDGTFSPSEAVDRGSAAKILAVVLGLPVDP 82

Qy   61 NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV 120
           :|:||||||.|.|:: |||||||||:..|||.. |.|..:|:||:.:|||| |:.|.|.:
Db   83 KAKPSFKDAQNHWAAPYIAAVEKAGVINGDGTGKFNPSSQINRASMASMLVQAYSLDKKI 142

RESULT 3
Q45759
ID   Q45759   PRELIMINARY;   PRT;   823 AA.
AC   Q45759;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Crystal protein.
GN   CRYTKD.
OS   Bacillus thuringiensis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1428;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TKD2-14;
RA   Hashimoto N.;
RT   "Bacillus thuringiensis subsp. mexicanensis TKD2-14.";
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; D86346; BAA13073.1; -.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF00395; SLH; 3.
SQ   SEQUENCE   823 AA;   87602 MW;   96FA83336A88ACE7 CRC64;

Query Match   16.8%;
Best Local Similarity   30.0%;   Pred. No. 6.4e-21;   DB 2;   Length 823;
Matches 272;   Conservative 105;   Mismatches 309;   Indels 222;   Gaps 37;

Qy   1 AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE 60
           |||||||| .|.|||.| |:|:|.||||||||||||.|||.|:|||||:||.|...||:|
Db   30 AGKSFPDVQPGSWSAEYIDYIVAKKAIEGKPDGTFAPTEAIDRASAAKIMAITIGLEVKE 89

Qy   61 NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV 120
           .|:||||||.:.|::||||||||||.|:||:.|.|.|..:|:|||.|:|:|..||:||.:
Db   90 GAKPSFKDADSWAAKYIAAVEKAGVIQGDETGKFNPNNQINRASMATMIVKAYKLDGKV 149

Qy   121 NGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYGK 180
           .|:| ..|.||.||||| |||||:.| |: ||..||:.||.||||||:|||||| .:||
Db   150 SGQLETKFSDLKDHWGEKAANILVALDITNGTGNGWEPDKSVTRAEAAKFIAKTDMQFGQ 209

Qy   181 KDNAQAYVTDVK----------VSBPT----KJILTGTGLDKLSADDVTLEGDKAV 222
           |.:| .::...| || || .: ||:||:| ||||||:|:.||
Db   210 K--AEAKVESIKAINAKEIEVKLGATVTEETVKNAKFTLS-QGVDNVTVVDAKIGEAKKS 266

Qy   223 AIEASTDG------TSAVVTL-----GGKVAPNKDLTVKVKNQ------SFVTKFVEV 264
           |:||:||| |||||| |||| |||:|||||.: |..|:::|
Db   267 VILTLEDGKRLEKNNAYVVTVEKLNTAEGKEIP-KALEVIFFNDEVAPTVSTVSTPDGKV 325

Qy   265 KKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVE---------YLNLANHD----VKFVANN 312
           | :|.:|| |.:.:.|.:||||:||:|| |:|| |||.||.| |||.::.
Db   326 KVVFSEKLD-----AAKTVTVIVNGQKVEGKVENNTFTSNQALNLQNGKEFSIIVTGATD 380

Qy   313 LDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLTVSNTGIITVKNLDTPAS 372
           ||.: .:: || :::||||.|||.|.||.|||.|:| :|.|:| ||.:...
Db   381 LADNTMEMYEG----------------KATYKVEKT------TAPEVKDVKVKEITAEGK 419

Qy   373 AIKNVVFALDADNDGVVNYGSKLSGKDPALNSQNLVGEKASLNKLVATIAGEDKVVDPG 432
           |: :|||:.|:||:|||:|.||:||||||||||||||||||||||| .:|:.|||||||
Db   420 ATLEVTFS-----------------BELATNGQGVVVVKGK-EELKSTIA--------- 452

Qy   433 SISIKSSNHGIISVVNNYITAEAAGEATLITIKVG--DVTKDVKFKVTTDSRKLVSVKANP 490
           .:.:. :.:|:|.|:.||:.||::||||||.|.|| :|:||.||||||||||||:|||.
Db   453 -VDAQDKTKAVIEVTGALKEKETATDLTVEF-VGYKDAANNVGSKVSK-----AVKVSK 504

Qy   491 D-------KLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVVTTD 543
           | ||.||.|| |:||| ||||:|| |::|:|||.|:|||||||.|||.|||||.
Db   505 DVVAPNVKVEADQNKAATFTF---------DKEV------TAQNGKLRVINLD 543

Qy   544 SSGIGTKTIGVTGNDVGEGTVHFQNGNGATL--------GSLVNVTEG--------NVAF- 588
           :|..|.|.|.||... .|||:|..|:|.: |||:.|| | |||:
Db   544 TSKDVTADVTVAPAK------DNKKGITLLTFPAKGNYKVAATKGLVKDAAGNESVAFT 595

Qy   589 KNFELVSKVGQYGQSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEGYEVES----KN 644

```
Qy  121 NGELVTTFEDLLDHWGEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTD----- 175
Db  143 IGELPTQFKDLEPHWGKKQANILVALEISKGTGNGWNPEGTVTRAEAAQFIAMADQNKTS 202
Qy  176 --KKYGKKDNAQAY-------VTDVKVSEPTKLITLGTGLDKLSADDVTLEGDKAVAIE 225
Db  203 TSKRMYMNRNVITHQPSLSSGITDVQ-HKPQMVEVT---------------------E 239
Qy  226 ASTDGTSAVVT-LGGKVAPNDLTVKVK--NQSFVTKFVYEVKKLAVEKL-TFDD----- 276
Db  240 QRADGWLKIVTSKGEKWTP---LTEKETINEEFTI--YETASHSSKVLGTYNAQTVTV 293
Qy  277 -DRAGQAIAFKLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIFEGGEATSTTGKLAV 335
Db  294 MEESGSWIRIRVGAGFQWVDKNQLNPVKQE-----NFLEGK-AIIIDPGHGGIDSGNVGY 347
Qy  336 GIKQGDYKVEVQV------TKRGGLTVSNTGIITVK-NLDTPASAIKNVVFALDADNDGV 388
Db  348 YEKESETVLDVSLRLKKIFEQKAPFTVMFTRTDNTRPGVNSTDSLKKRVEFAQEHNGDIF 407
Qy  389 VNYGSKLSGKDPALNSQNL---------VVGEKASLNKLVATIAGEDKVVDPGSISIKSSN 440
Db  408 VSIHANGSAEKNGQTETIYYQSARAKVTNPHVEDSKLLA-----QKIQDRIVAALGTKD 462
Qy  441 HGIISVVNNYITAEAAGEATLITIKVGDVTKDVKFKVTTDSRKLVSVKA 488
Db  463 RG-VKHQDLYVTRENTMPAVLTELAFVDNKSDADKIATPKRQAAAEA 509
```

RESULT 5
Q81S57
```
ID   Q81S57        PRELIMINARY;       PRT;    414 AA.
AC   Q81S57;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   N-acetylmuramoyl-L-alanine amidase, family 3.
GN   BA1817.
OS   Bacillus anthracis (strain Ames).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=198094;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22608414; PubMed=12721629;
RA   Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA   Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA   Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA   Kolonay J.F., Beanan M.J., Dodson R.J., Durkin A.S., Gwinn M.,
RA   DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA   Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA   Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA   Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA   Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA   Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA   Fraser C.M.;
RT   "The genome sequence of Bacillus anthracis Ames and comparison to
RT   closely related bacteria.";
RL   Nature 423:81-86(2003).
DR   EMBL; AE017029; AAP25724.1; -.
DR   TIGR; BA1817;
DR   GO; GO:0008745; F:N-acetylmuramoyl-L-alanine amidase activity; IEA.
DR   GO; GO:0009253; P:peptidoglycan catabolism; IEA.
DR   InterPro; IPR002508; Amidase_3.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF01520; Amidase_3; 1.
DR   Pfam; PF00395; SLH; 3.
DR   SMART; SM00646; Ami_3; 1.
KW   Complete proteome.
SQ   SEQUENCE   414 AA;  45351 MW;  40319D11B618F18F CRC64;

Query Match     13.2%;  Score 556.5;  DB 16;  Pred. No. 3.1e-15;  Length 414;
Best Local Similarity   39.2%;
Matches 145;  Conservative 54;  Mismatches 120;  Indels 51;  Gaps 14;
```

```
Qy    3 KSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENA 62
Db   27 KTFPDVQTGHWAEDSINYLAEKGAVTGNEKGMFEPGKEITRAPAATMAKILNLPIDKNA 86
Qy   63 QPSFKDAKNIWSSKYIAAVEVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNG 122
Db   87 KPSYADSQKEWATPIIAAVEVEKAGVVKGTG-NGFEPDGKIDRVSMASLLVEAYKLESKVNG 145
Qy  123 ELVTIFEDLLD-HWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDK-KYGK 180
Db  146 TPATKFKDLETLNWGKEKANILVELGISVGTGDKWEPKKILTKAEAAQFIKKADSLKVGN 205
Qy  181 KDNAQAYVTDV------KVSEPTKLTL-TGTGLDKLSADDVTLEGDKAVAI--EA 226
Db  206 PLVEKVVIIDPGHGGFDPGNPGQGVEESKIVFDTSLRLQKLLEKNTPL---KALLTREEN 262
Qy  227 STDGTSAVVTLGGKV-----APNKDLTVKV-----KNQSFVTKFVYEVKKLAVEKLTFDD 276
Db  263 GNPGSNKNESLANRVKFGQENNADIFVSIHANSSEKHDGHGTETYYYKKSKRGETQIEK 322
Qy  277 DRAGQAIAFKLNDEKGNADVEYLNLANHDVK------FVANNLDGSPA-------NIFE 322
Db  323 D--SEVLAKKIQ----NRVVEALHTRDRKIKDDHSLYVVNN-NTVPAVLTELAFIDNDID 375
Qy  323 GGEATSTTGK 332
Db  376 NGKLATESGR 385
```

RESULT 6
Q813K8
```
ID   Q813K8        PRELIMINARY;       PRT;    483 AA.
AC   Q813K8;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   S-layer protein / peptidoglycan endo-beta-N-acetylglucosaminidase
DE   (EC 3.2.1.-).
GN   BC2464.
OS   Bacillus cereus (strain ATCC 14579 / DSM 31).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=226900;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22608411; PubMed=12721630;
RA   Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA   Kapatral V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA   Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,
RA   Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA   Overbeek R., Kyrpides N.;
RT   "Genome sequence of Bacillus cereus and comparative analysis with
RT   Bacillus anthracis.";
RL   Nature 423:87-91(2003).
DR   EMBL; AE017006; AAP09426.1; -.
DR   GO; GO:0004040; F:amidase activity; IEA.
DR   GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   GO; GO:0009253; P:peptidoglycan catabolism; IEA.
DR   InterPro; IPR002301; Amidase_4.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF01832; Amidase_4; 1.
DR   Pfam; PF00395; SLH; 3.
DR   SMART; SM00047; LYZ2; 1.
KW   Glycosidase; Hydrolase; Complete proteome.
SQ   SEQUENCE   483 AA;  53801 MW;  13D0CCA1F2911A6D CRC64;

Query Match     12.6%;  Score 528.5;  DB 16;  Length 483;
Best Local Similarity   31.3%;  Pred. No. 5.1e-14;
Matches 147;  Conservative 70;  Mismatches 145;  Indels 107;  Gaps 14;

Qy    5 FPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQP 64
```

Db   40 FPDVPT--WADKSVNYLVLVDKQVLNGYPDGTFGSNDSLDRASATKIMTKVLGIKIDPNAKP   97
           |||||| : ||| : ||||| | : : | ||||| ||||||| : ||||: : ||| ||
Qy   65 SFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGEL   124

Db   98 SFTDSQNHWATPYIAAEKAGIVKGEGNGIFNPSGKVTRAMAATMLVNAYKLQSTANHNE   157
           ||:|| ||| : |||| |||| || : || |||| : : ||| | || || :
Qy   125 VTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYGKKDNA   184

Db   158 QVRFEDLKGHWGEKHANILILDLKISNGTENGWQPNRFITRAERAAQLTAKTDMLQQNLNDE   217
           : | ||| || ||
Qy   185 QAYVT-----DVKVSEPTKLT------LTGTGLDKLSADD----VTLEGDK   220

Db   218 KEVITATSYEDLNLTVASKITAQEIDSFIAKYHSDSPLVGHGQDFINAQNQYGVSAHYLA   277
           : : |:
Qy   221 AVAIEASTDGTSAVV--------------------TLGGKVAPNKD--------   246

Db   278 AHAILESGYGKSEIAYQKHNLFGLRAYDGDPFKYAKYLPSYGDSIAYNANYVRERYLEBS   337
           : :|
Qy   247 --------LTVKVKNQSFVTK----FVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEK   291

Db   338 GMYYNGSTLTGMNVKYASDXGWAKKIAGIMERIKPFRVEDYTY-------AKKLPKNP   388
           : :
Qy   292 GNADVEYLNLANHDVKFVANNLDGSPANIFBGGEATSTTGKLAVGIKQGDYKVEVQVTKR   351

Db   389 ETLDVDAL---SNEIPY-KMYADGSSSNV----VSSAT------YYQVPFPFNLKIKSR   433
           : ::
Qy   352 GGHTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDF   400

Db   434 PDVAVEENKVGTV-----TPGTTI----FIYREDPNGWVEFSFEANGEKY   474

RESULT 7
Q8KYE8
ID   Q8KYE8   PRELIMINARY;   PRT;   531 AA.
AC   Q8KYE8;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Surface-layer N-acetylmuramoyl-L-alanine amidase. (pXO2-42).
GN   BXB0045.
OS   Bacillus anthracis.
OG   Plasmid pXO2.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1392;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A2012;
RX   MEDLINE=22061436; PubMed=12004073;
RA   Read T.D., Salzberg S.L., Pop M., Shumway M., Umayam L., Jiang L.,
RA   Holtzapple E., Busch J.D., Smith K.L., Schupp J.M., Solomon D.,
RA   Keim P., Fraser C.M.;
RT   "Comparative Genome Sequencing for Discovery of Novel Polymorphisms in
RT   Bacillus anthracis.";
RL   Science 296:2028-2033(2002).
DR   EMBL; AE011191; AAM26204.1; -.
DR   GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR   GO; GO:0008745; F:N-acetylmuramoyl-L-alanine amidase activity; IEA.
DR   GO; GO:0009253; P:peptidoglycan catabolism; IEA.
DR   InterPro; IPR002508; Amidase_3.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF01520; Amidase_3; 1.
DR   Pfam; PF00395; SLH; 3.
DR   SMART; SM00646; Ami_3; 1.
KW   Plasmid.
SQ   SEQUENCE   531 AA;   58860 MW;   46CBBFDD398DD7AC CRC64;

Query Match   12.6%; Score 528; DB 2; Length 531;
Best Local Similarity   33.2%; Pred. No. 6.1e-14;
Matches 158; Conservative 68; Mismatches 184; Indels 66; Gaps 17;

Qy   3 KSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENA   62
           | : | | ||:|:|| : |: |||: | ||:|||||:||||:| | ||| :||
Db   31 KTFTDVP--NWAQQSVNYLMKK-ALDGKPDGTFSPSEKIDRGSAAKLMAMVLGLQINKQA   87

Db   63 QPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNG   122
           : ||| ||||||||||||||||||||||| ||| ||||||||||||||||||||||||| :::
Qy   88 KPSFQDAKNHWASPYIIAAVEKAGVIYGDGSGNFNPSKDIDRASMASMLVEAYKLNNRIIG   147

Qy   123 ELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYGKKD   182
           : || ||||||| || ||::: | ::|| :: ||| | | | |||:|||||| : : : :::
Db   148 DLPTQFEDLKGHWGAKLANALVALGISKGTDGWKPNGIVTRAEAVQFIAQTDMK--KAD   205

Qy   183 NAQAVTD--VKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVT-LG   238
           : ||| :: | : ||::|:||: : : |||||| : ||| :|| :: :|
Db   206 TSKRMYMNRHFITYHQP---SLSSGVTSNQHAPQI------IVKEQRADGWIKIVTNIG   256

Qy   239 GKVAP--NKDLTVKVKNQSFVT--KFVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEKG--   292
           | | || : |:: ||| ::: ||||||||||||||||||||||| :
Db   257 DKWTPLYEKRETI---HSTFTTYPEASHSSKVLGT-------HSPQTVT--VIEEKGSW   303

Qy   293 --NADVEYLNLANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGD--YKVE   345
           : :: : |||| | | ||||||||||||||||||||||||||||
Db   304 IRIRTNAGFQWLDKNQLTLPKKQNNFLEGKTIIIDPGHGGIDGGHKGIYMNESPVVYDTA   363

Qy   346 VQVTKRGGLTVSNTGIITVKNLDTPA----SAIKNVVFALDADNDGVVN--------   390
           :: |
Db   364 VRVQKLFAQKTPFTALLTRDAYSRPGKNATDSLGKRVEFAKKNKGDIFVSIHANGFNGNA   423

Qy   391 ------YGSKLSGKDFALNSQNLVVGEKASLNKLVATIAGEDKVVDPGSISIKSSN   440
           : : |:
Db   424 HGTETFYYKAPTQKSNPVYNDSRILAEKIQ-KRIITALQTRDRGVKIGNLYVLREN   478

RESULT 8
Q81U01
ID   Q81U01   PRELIMINARY;   PRT;   344 AA.
AC   Q81U01;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   S-layer protein, putative.
GN   BA1093.
OS   Bacillus anthracis (strain Ames).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=198094;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22608414; PubMed=12721629;
RA   Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA   Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA   Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA   Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA   DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA   Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA   Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA   Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA   Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA   Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA   Fraser C.M.;
RT   "The genome sequence of Bacillus anthracis Ames and comparison to
RT   closely related bacteria.";
RL   Nature 423:81-86(2003).
DR   EMBL; AE017027; AAP25072.1; -.
DR   TIGR; BA1093; -.
DR   InterPro; IPR006635; NEA_transpt.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF05031; NEAT; 1.
DR   Pfam; PF00395; SLH; 3.
DR   SMART; SM00725; NEAT; 1.
KW   Complete proteome.
SQ   SEQUENCE   344 AA;   38277 MW;   48764AC55FD72241 CRC64;

Query Match   12.1%; Score 509; DB 16; Length 344;
Best Local Similarity   59.9%; Pred. No. 2e-13;
Matches 103; Conservative 22; Mismatches 45; Indels   2; Gaps   1;

```
Qy    5  FPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQP  64
            ||| ::| ||||| || ||||||||  ||||||||:||| ::||:||
Db   169  FSDVPT--WAQESVCYLVDKEAVHGKPDGTFAPAESIDRSSAAKILA:VLRLEIKKDAKP 226

Qy   65  SFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGEL 124
            |||||  | | ||| || ||||||||||||||| |||| |||
Db   227  SFPDAQNHWATPYIAAVEKAGIVKGDEKGNFNPSGLINRASMASMLVNAYKLERNENIKL 286

Qy  125  VTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDK 176
            || |||| || : | |||||||||||   ||  :||| |||
Db  287  PKEFADLNNHWGAKYANILIIQEKISIGTDNGWAPNKAVSRAERAQFIAKADK 338

RESULT 9
Q81QA4
ID   Q81QA4;
AC   Q81QA4;
DT   01-JUN-2003  (TrEMBLrel. 24, Created)
DT   01-JUN-2003  (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003  (TrEMBLrel. 25, Last annotation update)
DE   N-acetylmuramoyl-L-alanine amidase, family 3.
GN   BA2528.
OS   Bacillus anthracis (strain Ames).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=198094;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22608414; PubMed=12721629;
RA   Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA   Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA   Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA   Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA   DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA   Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA   Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA   Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA   Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA   Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA   Fraser C.M.;
RT   "The genome sequence of Bacillus anthracis Ames and comparison to
RT   closely related bacteria.";
RL   Nature 423:81-86(2003).
DR   EMBL; AE017032; AAP26383.1; -.
DR   TIGR; BA2528; -.
DR   GO; GO:0008745; F:N-acetylmuramoyl-L-alanine amidase activity; IEA.
DR   GO; GO:0009253; P:peptidoglycan catabolism; IEA.
DR   InterPro; IPR002508; Amidase_3.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF01520; Amidase_3; 1.
DR   Pfam; PF00395; SLH; 3.
DR   SMART; SM00646; Ami_3; 1.
KW   Complete proteome.
SQ   SEQUENCE  410 AA;  45230 MW;  28F7F51302F92428 CRC64;

Query Match         12.0%;  Score 506;  DB 16;  Length 410;
Best Local Similarity  34.1%;  Pred. No. 3.4e-13;
Matches  141;  Conservative  62;  Mismatches 139;  Indels  72;  Gaps  15;

Qy    5  FPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQP  64
            ||| :: | ||  || |||||||||| |||||||:||  ::||:||
Db   29  FPDVPA--WADKSVTYLVDKQVLSGYPDGTFGSSDTLDRASAATIMTKALGIHIDLNAKP  86

Qy   65  SFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGEL 124
            |||||  | || || || |||||||  |||| || |||| |||
Db   87  SFKDSQNHWGTPYIAAEKAGIIKGEGNGIFNPSGKVTRAAMATMLVNAYKLQNKNTSNG  146

Qy  125  VTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYGKKDNA 184
            :: || |||| || : | |||||||||||  || :||| |||
Db  147  QSKFEDLKGHWGEKFANTLIDLKISVGTDNGWCPNKFTTRAERAQLTAKTDMLQYSHSNP 206

Qy  185  QAYVT--------DVKVSEPTKLTL-TGTGLDKLSADDVTLEGDKAVAIFASTD 229
                                  : ||  || : ||
Db  207  LENKTIIIDPGHGGEDPGKDTKGLPESKIVLDTSLRLQKL----------LEKHTP 252
```

```
Qy  230  GTSAVTLGGKVAPNKD----LTVKVK----NQSFVTKFVYEVKKLAVEKLTFDDDRAGQ  281
            ::|                 : :|              ::  | ::| |||
Db  253  FT-VLLTRKSDTRPGHDQKSSLQERVKFAKQNQGDIP------ISVHANAFNGNAKGT  303

Qy  282  AIAF--KLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQ 339
            :|:|    :   |: | :  : : :|        :|  :::
Db  304  ETYYYKSSKSEKTNPHVE------ESRVLABKIQTRLV-----DALQTRDR--GVKH  347

Qy  340  GDYKVEVQVTKRGGLTVSNTGIITVKN-LDTPASAIKN-VVPALDADNDGVVNY 391
            ||  :|        | :| |||  |  ::|    : |  ::|  ||  |
Db  348  GD----LHVIRENDMPAVLTELAFIDNGIDYSKLSTENGRQIAAEAIYEGILDY 397
```

RESULT 10
```
Q81AN7
ID   Q81AN7;                     PRELIMINARY;     PRT;   410 AA.
AC   Q81AN7;
DT   01-JUN-2003  (TrEMBLrel. 24, Created)
DT   01-JUN-2003  (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003  (TrEMBLrel. 25, Last annotation update)
DE   S-layer protein / N-acetylmuramoyl-L-alanine amidase (EC
DE   3.5.1.28).
GN   BC3524.
OS   Bacillus cereus (strain ATCC 14579 / DSM 31).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=226900;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22608415; PubMed=12721630;
RA   Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA   Kapatral V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA   Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,
RA   Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA   Overbeek R., Kyrpides N.;
RT   "Genome sequence of Bacillus cereus and comparative analysis with
RT   Bacillus anthracis.";
RL   Nature 423:87-91(2003).
DR   EMBL; AE017009; AAP10458.1; -.
DR   GO; GO:0016787; F:hydrolase activity; IEA.
DR   GO; GO:0008745; F:N-acetylmuramoyl-L-alanine amidase activity; IEA.
DR   GO; GO:0009253; P:peptidoglycan catabolism; IEA.
DR   InterPro; IPR002508; Amidase_3.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF01520; Amidase_3; 1.
DR   Pfam; PF00395; SLH; 3.
DR   SMART; SM00646; Ami_3; 1.
KW   Hydrolase; Complete proteome.
SQ   SEQUENCE  410 AA;  45571 MW;  318702F62DE6341 CRC64;

Query Match         12.0%;  Score 504;  DB 16;  Length 410;
Best Local Similarity  54.7%;  Pred. No. 4.1e-13;
Matches  98;  Conservative  27;  Mismatches  52;  Indels  2;  Gaps  1;

Qy    5  FPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQP  64
            |||||  :   ||||||||:|||||||| |||||||||| |||
Db   29  FPDVPA--WADKSVNYLVDKQVLNGYPDGTFGSNDSLDRASATKIMTKVLGLQIDSNAKP  86

Qy   65  SFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGEL 124
            || ||| ||| || || ||:|| || || |||||| ||| :|||
Db   87  SFTDSQNHWATPYIAAEKAGIVKGEGNGIFNPSGKVTRAAMATMLVNAYKLQSTPNHNE 146

Qy  125  VTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYGKKDN 183
            || |||| || : | |||||||||||  || :||| |||
Db  147  QVKFEDLKGHWGEKYANILIGLKISNGTENGWQPNRFITRAEAAQLTAKTDMMQHRQKN 205

RESULT 11
Q81HB4
ID   Q81HB4;                     PRELIMINARY;     PRT;   530 AA.
AC   Q81HB4;
DT   01-JUN-2003  (TrEMBLrel. 24, Created)
DT   01-JUN-2003  (TrEMBLrel. 24, Last sequence update)
```

```
QSL655
ID   Q9L655          PRELIMINARY;      PRT;    931 AA.
AC   Q9L655;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Surface layer protein.
GN   SLPA.
OS   Bacillus pseudofirmus.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=79885;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OF4;
RX   MEDLINE=20485546; PubMed=11029415;
RA   Gilmour R., Messner P., Guffanti A.A., Kent R., Scheberl A.,
RA   Kendrick N., Krulwich T.A.;
RT   "Two-Dimensional Gel Electrophoresis Analyses of pH-Dependent Protein
RT   Expression in Facultatively Alkaliphilic Bacillus pseudofirmus OF4
RT   Lead to Characterization of an S-Layer Protein with a Role in
RT   Alkaliphily.";
RL   J. Bacteriol. 182:5969-5981(2000).
EMBL; AF242295; AAF68436.1; -.
InterPro; IPR001119; SLH.
Pfam; PF00395; SLH; 3.
SQ   SEQUENCE 931 AA; 96855 MW; 6A9727171C0A78D0 CRC64;

Query Match             9.4%;
Best Local Similarity   24.2%;        Pred. No. 3.2e-08;   Score 395;  DB 2;   Length 931;
Matches 233;   Conservative 106;   Mismatches 336;   Indels 288;   Gaps 43;

Qy     1 AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVD-  59
Db    34 ADAKFSDVSSSHWAIDDINYLVEKGAIQGYPDGTFKPNNSITRAEIAVVLANTLGLDVDS  93

Qy    60 ENAQPSFKDA-KNIWSSKYIAAV--EKAGVVKGDGKEN-FYPEGKIDRASFASMLVSAY  114
Db    94 EVTTDKFSDVPATHWANPYIAAIVDQTEGVI--DGYENGTFRPSNTITRQEMAKMVVEAY 151

Qy   115 NLKDKVNGELV---TTFEDLLDHWGEEKANILINLGISVG-TGGKWEPNKSVSRABAAAQ 169
Db   152 DL------ELVEGKDLPFTDVSGLWSTDYINILASNGVAAGMTATTFAPRGEVLRAQTAA 205

Qy   170 FI--ALTDKKYGKKDNAQAYVTDVKVSEPTKLTLT----------GTGLDKLSADDVTL  216
Db   206 FIHRARVEEERIEVPGLVPTVSSVTAVNATTLTVVLSDDTTHBVTLDTALVANEATEVTF 265

Qy   217 E-----------------------GDKAVAIEASTDGTSAVVTLGGKVAPNKDLTVKVKNQSFVTKFVY 262
Db   266 EINDVEYTETVTWVYEELAVASVNAVNGKQLVVNFTQPVDASSVLTSPTASTSVKTGVL 325

Qy   263 EVKKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIFE 322
Db   326 SLNR------------TSTDTHSNDVV-----------LGVAERAVL----SSDGKTLTV- 358

Qy   323 GGEATSTTGKLAVGIKQGDYKVEVQ--VTKRGGLTVSNTGIITVKNLDTPASAIKNVVFALD 382
Db   359 ----TAPTGQFF----KGNYDVTVKDAKSGQNTIPSV-IKTISVDDTTAPIVNSVVYVPA 409

Qy   383 ADNDGVV--NYGSKLSGKDFALNSQNLVVGEKA---SLNKL----------IA  423
Db   410 TDKFEVTLSEPIDSLTGEVLRINGQVASGFDALTGPTNKLTPARPSSVATGTNATIYIA 469

Qy   424 G------------------------EDKVVDPGGSISIKSSNHG 442
Db   470 GFSDAAGNFVTPSTTTVPVTQDTTALAVASLEQVTNQKVRLTFNKELNSASKTALESGTG 529

Qy   443 II------SVNNY-----ITABAAGEATLIKVGDVT----KDVKFKVT 477
Db   530 VVVTRPNGSTTSNFTVAQNLIVDNTGK-TYDITLSDATYTNGNSEVFGITILIKDAFTDVT 588

Qy   478 TDSRKLVS-------------------VKANPDKLQVVQNKTLPVTFVTTDQYGDPFGANT 519
```

```
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   S-layer protein / N-acetylmuramoyl-L-alanine amidase (EC
DE   3.5.1.28).
GN   BC0902.
OS   Bacillus cereus (strain ATCC 14579 / DSM 31).
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=226900;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22608415; PubMed=12721630;
RA   Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA   Kapatral V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA   Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,
RA   Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA   Overbeek R., Kyrpides N.;
RT   "Genome sequence of Bacillus cereus and comparative analysis with
RT   Bacillus anthracis.";
RL   Nature 423:87-91(2003).
EMBL; AE017000; AAP07889.1; -.
GO; GO:0016787; F:hydrolase activity; IEA.
GO; GO:0008745; F:N-acetylmuramoyl-L-alanine amidase activity; IEA.
GO; GO:0009253; P:peptidoglycan catabolism; IEA.
InterPro; IPR002508; Amidase_3.
InterPro; IPR001119; SLH.
Pfam; PF01520; Amidase_3; 1.
Pfam; PF00395; SLH; 3.
SMART; SM00646; Ami_3; 1.
PROSITE; PS01072; SLH_DOMAIN; 1.
KW   Hydrolase; Complete proteome.
SQ   SEQUENCE 530 AA; 58834 MW; 37E448FB109D2598 CRC64;

Query Match             11.3%;
Best Local Similarity   30.3%;        Pred. No. 8.7e-12;   Score 475;  DB 16;   Length 530;
Matches  150;   Conservative   79;   Mismatches  210;   Indels  56;   Gaps  19;

Qy     3 KSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENA  62
Db    27 KKFPDVSDSAWSKDAIYLVERNVINGMPYDGNFMPYGNLTRAQAAKIIATAIGAKVDPNA  86

Qy    63 QPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNG 122
Db    87 KPSYNDAKNSWAASFIAAMEKENIIKGREPGVPDPEGKVTRAEMAAMLVRAYNLKSKVTG 146

Qy   123 ELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAAQFIALTDKKYGKKD 182
Db   147 PVPTKFADLENHWGKEEVNILVELKLSLGTENGWKPNDSIIREQRAQLTAQTD-KFSK-- 203

Qy   183 NAQAYVTDVKVSEPTK-LTLTGTGLDK-LSADDVTLEGDKAVAIEASTDGTSAVVTLGG- 239
Db   204 NSDRPVETKKMYIDRKFITYHAPSLSSGISANQ---HNPQTVEIKEBERDGWIKIATSNGD 260

Qy   240 KVAPNKDLTVKVKNQSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAFKLNDE---KGNAD  295
Db   261 KWTPLVEKT-EVINEGFTT-----YARASSSKVMGTHNAQQVTIEENGSWIRIRMGAG   314

Qy   296 VEYINLANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQG----DYKVEVQ--VT 349
Db   315 FQWVNKNQLNPVKQGNFLEGK-AIIIDPGHGGVDPGHSGVKMDESAIVLDTSLRVQKLFE 373

Qy   350 KRGGLTVSNTGIITVKNLDTPASAI-KNVVFALDADNDGVVN-----YGSKLSGKD--- 399
Db   374 QKTPFTVLLTRNDDTRPGNTPGESLKKKRVEFAQENKGDIFVSIHANGFNEQVEGTETFYY 433

Qy   400 -FALN---SQNLVVGEKASLNKLVATIAGEDKVDPGGISIKSSNHGIISVVN------- 448
Db   434 RSATNPNSEESRVLAEKVQ-KRLVQALQSNDR------GVKTENFYVKYNTMPAILAE 485

Qy   449 -NYITAEAGEATLT 462
Db   486 LGFIDAKGSGEKLAT 500

RESULT 12
```

```
Db   64 PSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGE  123
           ||  :  | ||   : || |||                       : |  |
Qy   91 AGFTDVPKD-RAKYVNALVEAGVLNGKAPGKFGAYDPLTRVEMAKIIANRYKLK---ADD  146

Db  124 LVTIPEDLLDHWGEKANLLINLGISVG-TGGKWEPNKSVSRAEAAQFIALTDKKYGKKD   182
         :     |       :||        | ||                  : :
Qy  147 VKLPFTDVNDTWA-PYVRALYKEVTKGKTPTSFGAYQNITRGDFAQFVY-------RAV   198

Db  183 NAQAVTDVKVSEPTKLILTGTGLDKLSADDVT-LEGDKAVAIEAST---DGTSAVVTLG   238
         :  |  : :       : | |   |   :  | :| |  | |
Qy  199 NINAVPEIVETAVNSTTVKVTFNTQIADVDFTNFAIDNGLTVTKATLSRDKKSVEVVVN   258

Db  239 GKVAPNKDLTVK---VKN------QSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAFKLND  289
         |    |:| |     ||        ||  | | |
Qy  259 KPFTRNQEYTITATGIKNLKGETAKELTGKFWSVQDAVTVALNNSSLKVGESSGLTVKD   318

Db  290 EKGNADVEYINLANHDVKFVANNLDGSPANI-FEGGEATSTTGKLAVGIKQGDYKVEVQ   347
               : :            :          |        ||    : | |
Qy  319 QDGK-----------DVVGAKVELTSSNTNIVVSSGEVSVSAAKV-TAVKPGTADVTAK   366

Db  348 VTKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNL  407
         || |  :| | ::  || |  |
Qy  367 VTLPDGVVLTNTFKVTVTEVPV---QVQNQGFTL-VDN------LSNA------PQNT   408

Db  408 VVGEKASLNKLVATIAGEDKVV-------DPGS------ISIKSSNHGIISVNN---    449
         |               :  :          |||        :||:| | |:|
Qy  409 VAFNKA--EKVTSMPAGETKTVAMYDTKNGDPETKPVDFKDATVRSLNPIIATAINGSE  466

Db  450 -YITAEA--AGEATLTIKVGDVTK------DVK------FKVTTDSRKLVSVKANPDKLQ  494
           :| |    |||:|| | || ||       ||       | |||| :|:| | |
Qy  467 LLVTANAGQSGKASPEVTFKDNTKRTFTVDVKDPVLQDIKVDATSVKLSDEAVGGGEVE  526

Db  495 VVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVABGGLDVVTTDSGSIGTKTIGV  554
          |:|| | | | :|||  |                          |  |  | |:|| |
Qy  527 GVNQKTIKVSAV--DQYG-----------KEI-KPGT--KGKVTVTTNTEGLVIKNVN-   568

Db  555 TGNDVGEGTVHFQNGNGATLGSLYV---NVTEGNVAFKNFELVS-----------     595
                   | | || |
Qy  569 -----SDNTIPFDSGNSATDQFVVVATKDKIVNGKVEVKYFKNASDTTPTSTKTITVNVV  623

Db  596 ------KVGQYGQSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEGYEVES-----K   643
               ||                    :   |:|||  || | :|| |:
Qy  624 NVKADATPVGLDIVAP-SEIDVNAPNTAS------TADVDFINFESVEIYTLDSNGNRLK  676

Db  644 NLAVADAKIVGNK--VVVTGK------------TPGKVDIHLTKNGAT------AGK    680
          | |||| ||| |    |  || ||
Qy  677 KVTPTATTLVGTNDVVEVNGNVLQFKGNDELTLLTSSSTVNVDVTADGITKRIPVKYINS  736

Db  737 ATVEIVQETIAIKSVNFKPVQTEN---FVEKKINIGTVLELEKSNLDDIVKGINLTKETQ  793
                                        :|:|              :::
Qy  737 ASVP-ASATVATSPVTVKLNSSDNDLTFEELIFGVIDFTQLVKD--EDINEFIAVSKAAK  793

Db  794 NDGYLINKPLVTVKDASGKVIPTGANVYGLNHDATNGNIWFDEEQAGLAKKPSDVHFDVD  853
          |  | :||   |       |      |    |        :      :   |
Qy  764 ---AGDVKLGDVTVSQTSDSALPNFKADLYDTLTTKYTDKG----TLVFK--VLKDKDV  813

Db  854 FSLANVVKTGSGTVSSS-------PSLSDAI--QLTNSGDAVSFTLVIKSIYVKGADK   902
         :|  |  | | || |||
Qy  814 ITSEIGSQAVHVNV  827

Db  903 DDNNLLAAPVSVNV  916
```

RESULT 13
Q45664
ID   Q45664   PRELIMINARY;   PRT;   920 AA.
AC   Q45664;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   SBSB gene (Sequence 5 from PATENT WO9906567 precursor).
GN   SBSB.
OS   Bacillus stearothermophilus.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX   NCBI_TaxID=1422;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PV72;
RX   MEDLINE=97197531; PubMed=9045827;
RA   Kuen B., Koch A., Asenbauer E., Sara M., Lubitz W.;
RT   "Molecular characterization of the Bacillus stearothermophilus PV72 S-
RT   layer gene sbsB induced by oxidative stress.";
RL   J. Bacteriol. 179:1664-1670(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PV72;
RA   Lubitz W., Resch S.;
RT   "SECRETION OF CARRIER-BONDED PROTEINS INTO THE PERIPLASMA AND THE
RT   EXTRACELLULAR SPACE.";
RL   Patent number WO9906567-A/5, 11-FEB-1999.
DR   EMBL; X98095; CAA66724.1; -.
DR   EMBL; AX000222; CAB77070.1; -.
DR   InterPro; IPR003343; Big_2.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF02368; Big_2; 2.
DR   Pfam; PF03568; Big_2; 2.
DR   SMART; SM00633; BID_2; 2.
DR   SMART; SM00635; SLH; 3.
KW   S-layer; Signal.
FT   CHAIN   32   920
SQ   SEQUENCE   920 AA; 97916 MW; 1F3C20344B40F3A2 CRC64;

Query Match   8.7%;
Best Local Similarity   23.3%;
Matches 227; Conservative 138; Mismatches 369; Indels 240; Gaps 48;
    Score 367; DB 2; Length 920;
    Pred No. 4.4e-07;

```
Qy    4 SFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQ  63
         ||:||      |:::|:|  ||| |||| | | | ::::|||||| : : ||:||:: | 
Db   33 SFTDVAPQY--KDAIDFLVSTGATKGKTETKFGVYDEITRLDAAVILARVLKLDVDNAKD  90
```

RESULT 14
Q9RER7
ID   Q9RER7   PRELIMINARY;   PRT;   1268 AA.
AC   Q9RER7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Surface layer protein precursor.

GN   SBPA.
OS   Bacillus sphaericus.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1421;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCM 2177;
RA   Ilk N., Egelseer E.M., Jarosch M., Sleytr U.B., Sara M.;
RT   "Nucleotid sequence of sbpA, the S-layer gene from Bacillus sphaericus
RT   CCM 2177.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF211170; AAF22978.1; -.
DR   HSSP; P22629; 1SWL.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF00395; SLH; 3.
DR   PROSITE; PS01072; SLH_DOMAIN; 2.
KW   Signal.
FT   SIGNAL   1   30   POTENTIAL.
SQ   SEQUENCE   1268 AA; 132046 MW; 2C51D40FADFD0886 CRC64;

Query Match   7.7%; Score 322; DB 2; Length 1268;
Best Local Similarity 23.8%; Pred. No. 4.5e-05;
Matches 197; Conservative 93; Mismatches 296; Indels 240; Gaps 36;

Qy   13 WAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQPSFKDA-KN   71
               ::|||: ||| ::: |: : : ::::||| : ||:::||:|:||::| :|
Db   42 YAKEAVQSLVDQGVIQGDTNGNFNPLNTVTRAQAAEIFTKALELEADGDV--NFSDVKKG   99

Qy   72 IWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGELVTTFEDL 131
               ::|||| :   ::| :| ::: |:|| : : :::|:| :|:||
Db   100 AWYYRSIAAVVANGIFEGVSANBFAPNKSLTRSEAAKVLVDAFGLE---GSESLSQFADA 156

Qy   132 --LDHWGEEKANILINLGISVGT--GGK--WEPNKSVSRAEAAQFIALTDKKYGKKDNAQ 185
             :::::   :   : :||| : ||| :::   :|:||| |: ||:: ::|
Db   157 SQVKGWAKSALETAVANGIFTGSEENGKLNLKPNAAITRQDFAVVFART-----LDLAV   210

Qy   186 AYV-TDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVTLGGKVAPN 244
               :|| ||: :: |: | | : :: | || :: : || :| :
Db   211 ETVDASVKAINNTTVEVT---FDE-------EVDNVEALKFKIEGLEV-----------   248

Qy   245 KDLTVKVKNQSFVTKFVYEVEVKKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLNLANH 304
               ||| :| : | |: : :| || ::   : :|:: | : |:
Db   249 KNAAVKQTN-----------KKVVV--------------------LTTEAQTADKEYV-- 275

Qy   305 DVKFVANNLDGSPANIFEGGEA--------------TSTTGKLAVGIKQGDYKVEVQTKRGGL 354
             ::: |:||| ::   :| ::| :| | :: | : |
Db   276 -----VTLDGEEIGKFKGIEAVVPKSIVLKTTNTQG--------------KVGNQVTLTADV 318

Qy   355 TVSNTGIITVKNLDTPASAI-KNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKA 413
               :: :|:: : |::: ||:   ::: :: | |:   :|:::|: :::
Db   319 GVKAAGIPVTFNVDAPTGSLNKDAVVEVYTNAEGIASY----SYTQYAPDADDVTV---- 370

Qy   414 SLNKLVATIAGEDKVVDPGSISIKSSNHGIIISVVNNYITABAAGEATLTIKVGDVTKDVK 473
             ::|:| || || :| |::| :::| :|:: ::::|||| || : :|| ::|::|
Db   371 ----YPTGAPLRAFGP----------VYWGVDNILTIB---EVTTGNTLANGVKKT- 410

Qy   474 FKVTTDSRKLVSVKANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVV- 532
             |:::|:   :| ||||||| ::|:: | ::   : :   :|::||: |||
Db   411 YKVTF-------KDPKTGAALTNKKLNVSFV----------ENTNVAFNAISKATVTN 451

Qy   533 AEGGLDV-----------VTTDSGSIGTKTIGVTGNDVGEGTVHFQNGNGATLGSLY 578
             :|| ||| |||| ::| ::: |||| :|||| :|:|| : || |:
Db   452 PSSGLTVTPYQTTTGLQEEIQVTTDSN--GQATFVVSGTNTAVTPYVFVDGSSSVLGVST 509

Qy   579 V-----NVTEGNVAFKNFELVSKVGQYGQSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPEN 634
             | :| :: |:| || :   : :|: |:: : |: ::| :|| :|
Db   510 VTGTNNVTQATTANKKWEATELTATAQVKFEGAQLNHQITVERDGEEEAAAQYGS---K 566

Qy   635 LEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVDIHLTKNGATAGKATVEIVQETIAIKS 694
             ||| ::|| || :| :||| :|| |:: :::: :|:| |::| : |:::| :
Db   567 LNGREYKVKVLD-KDGKPYANGLVNIG-----LDEVLDRNLNTNSKAQFANVTDGTAL-- 618

Qy   695 VNFKPVQTENFVEKKINIGTVLELEKSNLDDIVKGINLTKETQHKVRVVKSGAEQCKLYL 754

Db   619 -----------------TLNPGT------------------------------
                              |||
Db   755 DRNGDA--VFNAGDVKLG------DVTVSQTSDSAL-----PNPKA 787
             :: ||| :: :   :|:   ::   ::|||
Qy   633 NSKGEATVVLYGAKGEVGTPVVWIDQNTSQNNQSGVLEDGEPPFKA 678

RESULT 15
Q9RB35
ID   Q9RB35   PRELIMINARY;   PRT; 1047 AA.
AC   Q9RB35;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   S-layer protein precursor.
OS   Cytophaga sp. 'Jeang 1995'.
OC   Bacteria; Bacteroidetes; Sphingobacteria; Sphingobacteriales;
OC   Flexibacteraceae; Cytophaga.
OX   NCBI_TaxID=85173;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Jeang 1995;
RX   MEDLINE=95291197; PubMed=7539663;
RA   Jeang C.L., Lee Y.H., Chang L.W.;
RT   "Purification and characterization of a raw-starch digesting amylase
RT   from a soil bacterium--Cytophaga sp.";
RL   Biochem. Mol. Biol. Int. 35:549-557(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Jeang 1995;
RA   Jeang C.L., Liao T.W., Chiou S.Y., Kang P.L., Shieh T.Y.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF068060; AAF21259.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR003343; Big_2.
DR   InterPro; IPR000515; BPD_transp.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF02368; Big_2; 2.
DR   Pfam; PF00395; SLH; 3.
DR   SMART; SM00635; BID_2; 2.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; 1.
KW   Signal.
FT   SIGNAL   1   32   POTENTIAL.
FT   CHAIN   33   1047   S-LAYER PROTEIN.
SQ   SEQUENCE   1047 AA; 108718 MW; 4395402E9965295A CRC64;

Query Match   7.3%; Score 307.5; DB 2; Length 1047;
Best Local Similarity 20.8%; Pred. No. 0.00014;
Matches 222; Conservative 153; Mismatches 368; Indels 325; Gaps 49;

Qy   17 SINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQPSFKDAKNIWSSK   76
               ::|:|| ||||||||| ||: : ||:| : ::::|||:|::|: :|
Db   43 AVNYLIEKGIASGLNKDTFGVQEKIKRGDAAIMVLTRLGVYGKDAKDAGFKDI-NSRVAK 101

Qy   77 YIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGELVTTFEDLLDHWG 136
               :::|: : ::| | | | ||:   :: :   :| ::: |:: | :
Db   102 AVNPLKEAGIISGKTATEFKPDANITRGEMAIIVAKAYKIQAVASKEV--PFKDATGIYK 159

Qy   137 EEKANILINLGISVGTGGKWEPNKSVSRAEAAQFI---------ALT---DK----K 177
               | ::|:|| || | : :| : :::| :| || :|: ||
Db   160 DAVAALYTNDIASGATADARGTTADITRGQFAIFLYKAEMIYQTSPVITLKGDKDVSVE 219

Qy   178 YGK-----------------------KDNAQAYVTDVKVSEPTKLTLTGTGLDK------ 208
             ||| |:||| ||:::| :|:| | ||:|
Db   220 YGADYADAGATAVDNKTVTVEVKTEVKDAAGPVVAAVDTKKVVGKYTVYSAVDKAGNKSE 279

Qy   209 ------------------------------------LSADDVTLEGDKAVA-IEASTDG 230
                                                 |::||:| || : : :||
Db   280 ACRTVTVKAQELKIQSVSALNSKQATVTFTSAFQKALTSANFDINGGLTVSDVKFSSSDR 339

Qy   632 ------------------------------------SQQGKIKL 632
                                                :|||:|:|
Db   695 VNFKPVQTENFVEKKINIGTVLELEKSNLDDIVKGINLTKETQHKVRVVKSGAEQCKLYL 754

```
Qy    231 TSAVVTLGGKVAPNKDLTVKV------KNQSFVTK--FVYEVKKLAVEKLTFDDDRAGQ 281
              ||||     :: :: |            :         |   : :  |   :::|
Db    340 KSATVTVSSDFTRDQEYSISVRGVKDAAGKEYAETKGSFTWKVADGVTVALGSTSLEQGQ 399

Qy    282 AIAFKLNDEKG----NADVEYLNLANHDVKFV-ANNLDGSPANIFEGGEATSTTGKLAVG 336
              || | |        : :  || ::|       |                     |||
Db    400 TVGLTVKDQAGKDVKDAKVE---VTSYNTNLVTVTNPDKDPATV----KVTANSDKLA-- 450

Qy    337 IKQGDYKVEVQVTKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLS 396
              :: |||    |       ||    |        :  || :        :       ::
Db    451 ---GSTDVVVETTLPDGSKLANTFTVNVKEAST--SISNAGYSLEADLD---TLGDDMA 501

Qy    397 G----KDFALNSQNLVVG--------------------------EKASLNKLVATIAGE 425
              :     | ::::|                               :  | | |  |  |:
Db    502 NTAAFKAFATPTTSVIEGSGTYELAAFGETNGNPDTAHVNFTGATRVTSSNPIVATALLD 561

Qy    426 DKVVDPGSISIKSSNHGIISVVNNYITAEAAGEATLIKVGDVTKDVKFKVTTDSRKL-- 483
              |          | |    |                 : ||||  | | |||     :
Db    562 D-----TAGVGVVKV------TPLSAGTTTVTVTMKD---------------GSRKSFP 594

Qy    484 VSVKANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVABGGLD--VVT 541
              :: |:| |    : :     ||         | :       :  | ::  :  |
Db    595 ITVTATPVQKAIDLNAT---TVKLSDE-----NRNEAA----DAAGATFQNGVDGKVVK 641

Qy    542 TDSGSIGTKTIG-------VTGNDVGEG-TVHFQNGNGATLG-SLYVNVTEGNVA---FKNF 591
              |   |  |  |          |                |      || || |     |||
Db    642 VTSLDQFNKEINKAAGKVTVSSSTEGLITLYKSDATGTTLGDQLKAGANELDFAGAKTTNF 701

Qy    592 ELVSK-----------VGQYGQSPDTK--------------LDLNVSTT----------V 616
              ::::                 |||                  :          |
Db    702 AILAQKDKPVSNATVTVSYFAKATDTKPAVTKTVSVNVTDIDPNTLTSAIDVVAASEIDA 761

Qy    617 EYQLSK------YTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVDIH 670
               :|| |      : ||                 :  |   |     :   : :
Db    762 NYEVSQAPGSIDFSSTPVYALDSKGNRLEKLTGTATIAD-KSAADKFVEVDNA-NKLKFK 819

Qy    671 LTKNGAT--AGKATVEIVQETIAIKSVNFKPVQTEN--FVEKKINIGT-----VLELEKS 721
               :           :||  |      |   |   |:        : |          | |:
Db    820 VDNNALTYLRAKNTVDVVVEADGVTKV--LPIAYKNTAVVPNKAAVSTNAVSVKLKSDAP 877

Qy    722 N--LDDIVKG-----INLTKETQHKVRVVKSGAEQGKLY------LDRNGDAVFNAGDVK 768
               :  ::|         | |               |:|         |  :  ::|  |
Db    878 NISINDILFGKADAEQLVKDGTDIIGVLKTAPNGGYLYNKPLVSITGTDGKALATGVNVY 937

Qy    769 LGDVTVSQTSDSALPNFKADLYDTLTTKYTDKG---------------------------- 801
              ||                  : ::|
Db    938 GGNALTLDAN----TFSNDLLETLI-QGYSTTGLAADSFYVDYMVANANDASTIAGDTVT 991

Qy    802 ------TLVFKVLKDKDVITSEIG---------SCAVHVNVLNNP 831
                    ||:       : :|                :|  :|:  ::
Db    992 PVAGKASTTLVVVKGIYVKDAVLNVEGKAFADASATEKAAH-NLLSSP 1038
```

Search completed: April 7, 2004, 17:31:54
Job time : 57 secs

```
                    GenCore version 5.1.6

              Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:        April  7, 2004, 17:19:59 ; Search time 62 Seconds
                                 (without alignments)
                                 3796.162 Million cell updates/sec

Title:         US-09-844-281-1
Perfect score: 4202
Sequence:      1 AGKSFPDVPAGHWAEGSINY.........ITSEIGSQAVHVNVLNNPNL 833

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:      1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database   :     A_Geneseq 29Jan04:*
            1:  geneseqp1980s:*
            2:  geneseqp1990s:*
            3:  geneseqp2000s:*
            4:  geneseqp2001s:*
            5:  geneseqp2002s:*
            6:  geneseqp2003as:*
            7:  geneseqp2003bs:*
            8:  geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	DB	Length	DB ID	Description
1	4202	100.0	5	833	AAM47757	Aam47757 Mature EA
2	719	17.1	4	785	AAB84651	Aab84651 Amino aci
3	564.5	13.4	6	529	ABU17672	Abu17672 Protein e
4	556.5	13.2	6	414	ABU17636	Abu17636 Protein e
5	553.5	13.2	6	525	ABU18750	Abu18750 Protein e
6	459	10.9	6	379	ABU17384	Abu17384 Protein e
7	369	8.8	6	408	ABU18670	Abu18670 Protein e
8	369	8.8	2	1252	AAR80530	Aar80530 B. sphaer
9	369	8.7	3	921	AAW22863	Aaw22863 Bacillus
10	367.5	8.7	3	921	AAB10626	Aab10626 B. stearo
11	367	8.7	3	920	AAW93253	Aaw93253 B stearot
12	257.5	6.1	2	1338	AAR41731	Aar41731 High mole
13	257.5	6.1	2	1598	AAW30291	Aaw30291 Non-typea
14	252.5	6.0	2	941	AAR77395	Aar77395 Full leng
15	251.5	6.0	2	941	AAR07478	Aar07478 Cellulase
16	250.5	6.0	2	1601	AAW30292	Aaw30292 Non-typea
17	249.5	5.9	2	1529	AAR41732	Aar41732 High mole
18	249.5	5.9	2	2353	AAR99393	Aar99393 Haemophil
19	246	5.9	3	1180	AAB01845	Aab01845 Haemophil
20	246	5.9	3	1188	AAB01844	Aab01844 Haemophil
21	242	5.8	3	1221	AAB01825	Aab01825 Haemophil
22	242	5.8	3	1227	AAB01824	Aab01824 Haemophil
23	242	5.7	3	2411	AAB23860	Aab23860 Haemophil
24	235	5.6	3	1004	AAB01841	Aab01841 Haemophil
25	235	5.6	3	1010	AAB01840	Aab01840 Haemophil
26	235	5.6	6	2712	ABU39146	Abu39146 Protein e
27	234.5	5.6	4	2314	AAB69136	Aab69136 M. catarr
28	231	5.5	7	1268	ADC00996	Adc00996 Enterchae
29	230.5	5.5	6	2154	ABU38771	Abu38771 Protein e
30	230	5.5	4	1325	AAG98256	Aag98256 Escherich
31	229	5.4	3	969	AAB01827	Aab01827 Haemophil
32	229	5.4	3	975	AAB01826	Aab01826 Haemophil
33	229	5.4	3	1612	AAB47333	Abb47333 Listeria
34	227.5	5.4	3	1073	AAB01837	Aab01837 Haemophil
35	227.5	5.4	3	1079	AAB01836	Aab01836 Haemophil
36	227	5.4	3	2504	ADA34534	Ada34534 Acinetoba
37	224.5	5.3	6	1386	ABB82573	Abb82573 H. influe
38	224.5	5.3	6	1449	ABB82570	Abb82570 H. influe
39	223.5	5.3	6	852	ABU15220	Abu15220 Protein e
40	222.5	5.3	6	2053	AAB69135	Aab69135 M. catarr
41	222	5.3	2	1992	AAW04505	Aaw04505 Moraxella
42	222	5.3	4	1992	AAB69137	Aab69137 M. catarr
43	222	5.3	4	1992	AAB69133	Aab69133 M. catarr
44	222	5.3	4	2047	AAB69134	Aab69134 M. catarr
45	222	5.3	6	2265	ABU17199	Abu17199 Protein e

```
                              ALIGNMENTS

RESULT 1
AAM47757
ID   AAM47757 standard; protein; 833 AA.
XX
AC   AAM47757;
XX
DT   25-FEB-2002  (first entry)
XX
DE   Mature EA1 protein.
XX
KW   Anthrax; antibacterial; vaccine; EA1 antigen.
XX
OS   Bacillus anthracis.
XX
PN   WO200183561-A2.
XX
PD   08-NOV-2001.
XX
PF   30-APR-2001; 2001WO-US013648.
XX
PR   28-APR-2000; 2000US-0200505P.
XX
PA   (TETR-) TETRACORE LLC.
XX
PI   Mangold BL, Aldrich JL, O'brien TW;
XX
DR   WPI; 2002-055457/07.
XX
PT   Novel monoclonal antibody, useful for detecting B.anthracis, and for
PT   treating B.anthracis infection, is specifically reactive against Bacillus
PT   anthracis and is non-reactive with B.thuringinesis and B.cereus.
XX
PS   Claim 6; Fig 1; 27pp; English.
XX
CC   The present invention relates to a monoclonal antibody which is
CC   specifically reactive against Bacillus anthracis EA1 antigen. The present
CC   sequence is the mature EA1 protein from Bacillus anthracis. The EA1
CC   protein corresponds to the eag gene. The monoclonal antibody is highly
CC   specific for Bacillus anthracis, and can distinguish Bacillus anthracis
CC   from closely related non-pathogenic species. The present sequence is
CC   useful for producing a vaccine against Bacillus anthracis and the
CC   antibody is useful for treating, preventing or controlling Bacillus
CC   anthracis infection (anthrax)
XX
SQ   Sequence 833 AA;

  Query Match             100.0%;  Score 4202;  DB 5;  Length 833;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-238;
```

```
Matches    833;    Conservative    0;    Mismatches    0;    Indels    0;    Gaps    0;

Qy      1 AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE 60

Qy     61 NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV 120

Qy    121 NGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYGK 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 NGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYGK 180

Qy    181 KDNAQAYYTDVKVSEPTKLTLTGTGLDKLSADDVTLBGDKAVAIEASTDGTSAVVTLGGK 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 KDNAQAYYTDVKVSEPTKLTLTGTGLDKLSADDVTLBGDKAVAIEASTDGTSAVVTLGGK 240

Qy    241 VAPNKDLTVKVKNQSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLN 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 VAPNKDLTVKVKNQSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLN 300

Qy    301 LANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLTVSNTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 LANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLTVSNTG 360

Qy    361 IITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLNKLVA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 IITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLNKLVA 420

Qy    421 TIAGEDKVVDPGSISIKSSNHGIISVNNYITAEAAGEATLTIKVGDVTKDVKFKVTTDS 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TIAGEDKVVDPGSISIKSSNHGIISVNNYITAEAAGEATLTIKVGDVTKDVKFKVTTDS 480

Qy    481 RKLVSVKANPDKLQVVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVV 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 RKLVSVKANPDKLQVVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVV 540

Qy    541 TTDSGSIGTKTIGVTGNDVGEGTVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQY 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTDSGSIGTKTIGVTGNDVGEGTVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQY 600

Qy    601 GQSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GQSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVT 660

Qy    661 GKTPGKVDIHLTKNGATAGKATVEIVQETIAIKSVNFKPVQTENFVBKKINIGTVLELEK 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GKTPGKVDIHLTKNGATAGKATVEIVQETIAIKSVNFKPVQTENFVBKKINIGTVLELEK 720

Qy    721 SNLDDIVKGINLTKETQHKVRVVKSGAEQGKLYLDRNGDAVFNAGDVKLGDVTVSQTSDS 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 SNLDDIVKGINLTKETQHKVRVVKSGAEQGKLYLDRNGDAVFNAGDVKLGDVTVSQTSDS 780

Qy    781 ALPNFKADLYDTLJTKYTDKGTLVFKVLKDKDVITSEIGSQAVHVNVLNNPNL 833
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ALPNFKADLYDTLJTKYTDKGTLVFKVLKDKDVITSEIGSQAVHVNVLNNPNL 833

RESULT 2
AAB84651
ID   AAB84651   standard; protein; 785 AA.
XX
AC   AAB84651;
XX
DT   05-SEP-2001   (first entry)
XX
DE   Amino acid sequence of a bacterial surface array protein (SAP).
XX
KW   Surface array protein; SAP; bacterial detection.
XX
OS   Bacillus anthracis.
XX
```

```
FH   Key            Location/Qualifiers
FT   Misc-difference  1..2
FT                  /note= "the nucleotides encoding these residues are not
FT                         given"
XX
PN   WO200149823-A2.
XX
PD   12-JUL-2001.
XX
PF   04-JAN-2001; 2001WO-US000358.
XX
PR   06-JAN-2000; 2000US-0174901P.
XX
PA   (BIOS-) BIOSITE DIAGNOSTICS INC.
XX
PI   Lee BA, Flores BM, Valkirs GE;
XX
DR   WPI; 2001-418358/44.
DR   N-PSDB; AAH28330.
XX
PT   Novel methods and kits for detecting the presence of Bacillus anthracis
PT   in a test sample.
XX
PS   Claim 2; Page 59-60; 62pp; English.
XX
CC   The present sequence represents a bacterial surface array protein (SAP).
CC   SAP is used in the method of the invention. The specification describes a
CC   method for detecting the presence of Bacillus anthracis in a test sample.
CC   The method comprises contacting the sample with a capture reagent and
CC   detecting whether the a SAP is bound to the capture reagent, which is
CC   indicative of the presence of B. anthracis in the sample. The method is
CC   useful for detecting the presence or absence of B. anthracis in a test
CC   sample
XX
SQ   Sequence 785 AA;
```

```
Query Match             17.1%;    Score 719;  DB 4;  Length 785;
Best Local Similarity   29.6%;    Pred. No. 8e-34;
Matches 233;            Conservative 123;  Mismatches 289;  Indels 142;  Gaps 29;

Qy      1 AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE 60
          |||||||||||| :||||||:||||::||||:||||||:|||||||:||||| :||||
Db      1 AGTFPDVPADHWGIDSINYLVEKGAVKGNDKGMPEPGKELTRABAATMMAQILNLPIDK 60

Qy     61 NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV 120
          : |||| ::|||||::||| ||| |||||| :||::|  | |||:|::|  ||||||::
Db     61 DAKPBFADSGQGWTYPFIAAVEKAGVIKGTG-NGFEPNGKIDRVSMASLLVEAYKLDTKV 119

Qy    121 NGELVTTFEDLLD-HWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYG 179
          ||||||||||||| ||:|||||||||||||||||||  :: ||||::||||||||  :
Db    120 NGTPATKFKDLETLNWGKEKANILVELGISVGTGDQWEPKKTVTKAEAAQFIAKTDKQFG 179

Qy    180 ----KKDNAQAYVT---DVKVSEPTKLTLTGTGLDKLSADDVTL----EGDKAVAIEA- 226
              : |::||:  :  |||:||||:|||:|:||||:||: | :      :|:|||||:|
Db    180 TEAAKVESAKAVTTQKVEVPSK---------AVEKLTKEDIKVTNKANNDKVLVKEVTL 230

Qy    227 STDGTSAVVTLGGKVAPNKDLTVKV-----KNQSFVTKFVYEVKKLAVEKLTFDDDRAGQA 282
          |||::::||: ::|||:::|: :|      |  |||||||:|| ::  |||  :|: ||
Db    231 SEDKRSATVELYSNLAAKQYTTVDVNKVGKTEVAVGSLEAKTIEMADQTVVADEP---TA 287

Qy    283 IAFKLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDY 342
          :|||:|:|||:|||:  : :|| ::   ||:|||||    |||:  ::||||||  ||:
Db    288 LQFTVKDENGTEVV------SPEGIEFV-----TPA------AEKINAKGEITLLAKGTS 329

Qy    343 KVEVQVTKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFAL 402
          ::|  ::  |:| :|||||||::| |||||||||::||::::::|||::|:||||||||:|
Db    330 TTVKAVYKKDGKVVAESKEVKVSASGAAVASISNWTVAEQNKAD------FTSKDFKQ 381

Qy    403 NSQNLVVGEKASLNKLVATIAGEDKVVDPGSISIKSSNH-GIISVVNNYITAEARAGEAT 460
          |:::||||:||:||| ::  : :|         ::  : ||:   |:|::||:||||||
Db    382 NNK-VYEGDNAYVQ---VELKDQFNAVTTGKVEYESLNTEVAVVDKATGKTVLSAGKAP 437

Qy    461 LTIKVGD-----------------------VTKDVKFKVTTDSRKLVSVKANPDKLQVVQNKTL 501
```

CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 529 AA;

Query Match   13.4%; Score 564.5; DB 6; Length 529;
Best Local Similarity   32.6%; Pred. No. 5.7e-25;
Matches 172; Conservative 81; Mismatches 192; Indels 83; Gaps 18;

Qy   1 AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE 60
           ||:|| :| ::||:   :::| :::| | :|||:|||||||
Db   25 AAKKFSDVPT--WAQQSVDYLVGKKALDGKPDGTFSPSEAVDRGSAAKILAVVLGLPVDP 82

Qy   61 NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV 120
            :|| |||:||| || ||||||||| | ||||:| ||| || ||||| ||||||
Db   83 KAKPSFKDAQNHWAAPYIAAVEKAGVINGDGTGKFNPSSQINRASMASMLVQAYSLDKKI 142

Qy   121 NGELVTTPEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQPIALTD----- 175
           :|||:|| |:|| ||||:||| |:||| ||:|| :|:| :|||||| |:|
Db   143 IGELPTQFKDLEPHWGKKQANILVALEISKGTGNGWNPEGTVTRAEAAQFIAWADQNKTS 202

Qy   176 --KKYGKKDNAQAY-------VTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIE 225
             :|| |:|||:   : |: ||:||| ||| :-E 239
Db   203 TSKRMYMNRNVITYHQPSLSSGITDVQ-HKPQWEVT-------------------

Qy   226 ASTDGTSAVVT-LGGKVAPNKDLTVKVK--NQSFVTKFVYEVKKLAVEKL-TFDD---- 276
            ::| || ||||||||||| |||| |||| || :|::| | | |:||
Db   240 QRADGWLKIVTSKGEKWTP---LTEKTETINEEFTT---YETASHSSKVLGTYNAQTVTV 293

Qy   277 -DRAGQAIAFKLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIFEGGEATSTTGKLAV 335
            :   :: ||::||| : |:||||:|||:|:||| :|
Db   294 MEESGSWIRIRVGAGFQWVDWDKGQLNFVKQE----NFLEGK-AIIIDPGHGGIDSGNVGY 347

Qy   336 GIKQGDYKVEVQV------TKRGGLTVSNTGIITVK-NLDTPASAIKNVFALDADNDGV 388
           ||:|||:|:|:| |||| |||||||:|||| |||||||:||||| ||:|:|||
Db   348 YEKESETVLDVSLRLKKIFEQKAPFTVMFTRTDNTRPGVNSTDSLKKRVEFAQEHNGDIF 407

Qy   389 VNYGSKLSGKDFALNSQNL----------VVGEKASLNKLVATIAGEDKVVDPGSISKSSN 440
           |: | ||| |||:| ::| | : ||: |::||||
Db   408 VSIHANGSAEKNGQGTETLYYQSARAKVTNPHVEDSKLLA----QKIQDRLVAALGTKD 462

Qy   441 HGIISVVNNYITAEAAGEATITKVGDVTKDVKFKFVTTDSRKLVSVKA 488
           :| |||:|: :::   : :::   ::| | ::|: | |
Db   463 RG-VKHQDLYVTRENTMPAVLTELAFVDNKSDADKIATPKKQRQAAAEA 509

RESULT 3
ABU17672
ID   ABU17672 standard; protein; 529 AA.
XX
AC   ABU17672;
XX
DT   19-JUN-2003 (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #3199.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Bacillus anthracis.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-0007285I.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,   Zamudio C,   Malone C,   Haselbeck R, Ohlsen KL, Zyskind JW;
PI   Wall D,   Trawick JD, Carr GJ,   Yamamoto R,   Forsyth RA, Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA21542.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 45596; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular

Db   438 VKVTVRDSKGKALVSHTVEIEAFAQKAMKDIKLEKTNVA---LSTKDVTD-------L 485
Qy   502 PVTFVTTDQYGDPFGANTAAIKEVLPKTG-VUAEGGLDVVTTDSGSGIGTKTIGVTGNDVG 560
Db   486 KVKAPVLDQYGKEFTAPVTV--KVLDKDGKELKEQKLEAKYVNRELV----LNAAGQEAG 539
Qy   561 EGTVHF--QNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQYGQSPDTKLDLNVSTTVEY 618
Db   540 NYTVVLTAKSGEKEAKATLALEL-KAPGAFSKFEV-------RGLDTELD-------- 581
Qy   619 QLSKYTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVD--------- 668
Db   582 ---KYVTEENQRNAMTVSVLPVDANGIVLKGAEAAELKVTTNKEGKEVDATDAQVTVQN 638
Qy   669 --IHLTKNGATAGKA-TVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLELEKSNLDD 725
Db   639 NSVITVGQGAKAGETYKVTKVTVVLDGKLITTHSFKVVDTAPTAK----GLAVEFTTSLKE 693
Qy   726 IVKGINL 732
Db   694 VAPNADL 700

RESULT 4
ABU17636
ID   ABU17636 standard; protein; 414 AA.
XX
AC   ABU17636;
XX
DT   19-JUN-2003 (first entry)

```
Db    27  KTFPDVQGTGHWAEDSINYLAEKGAVTGNEKGMPEPGKEITRAEAATMMAKILNLPIDKNA  86
                 ||||||||  |||||| || || |||| ||  |  |||| |:   ||: :|||| :||
Qy    63  QPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNG 122
                 :||: ||:   :::  ||||||||||||||  ||  |  |||||  |:|||||||||
Db    87  KPSYADSQKHWATPIIAAVEKAGVVKGTG-NGFEPDGKIDRVSMASLLVEAYKLESKVNG 145

Qy   123  ELVTTFEDILD-HWGEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDK-KYGK 180
                 |:|||  |||  |||||||||||  |||| ||: ||||  | |::  ||||  |||||
Db   146  TPATKFKDLETLNWGKEANILVELGISVGTGDKWEPKKILTKAEAAQFIKKADSLKVGN 205

Qy   181  KDNAQAYVTDV----------KVSEPTKLTL-TGTGLDKLSADDVTLBGDKAVAI--EA 226
                 :    |: :            :||: |:||  |:||||  |:|
Db   206  PLVEKVVIIDPGHGGFDPGNPGQGVEESKIVFDTSLRLQKLLEKNTPL--KALLTREEN 262

Qy   227  STDGTSAVVTLGGKV-----APNKDLTVKV------KNQSFVTKFVYEVKKLAVEKLTFDD 276
                 ||    |  |::  |       | |: |:|:      |::    |:  |:  ||||||||
Db   263  GNPGSNKNESLANRVKFGQENNADIFVSIHANSSEKHDGHGTETYYYKKSKRGETQIEK 322

Qy   277  DRAGGAIAFKLNDEKGNADVEYLNLANHDVK----FVANNLDGSPA----------NIFE 322
                 |   |  |||  :|| :  ||  :|||     ||    |  :  |  |:|
Db   323  D--SEVLAKKIQ----NRVVEALHTRDRKIKDDHSLYVVNN-NTVPAVLTELAFIDNDID 375

Qy   323  GGEATSTTGK 332
                 |:  :|:|::
Db   376  NGKLATESGR 385
```

```
RESULT 5
ABU18750
ID   ABU18750 standard; protein; 525 AA.
XX
AC   ABU18750;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #4277.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
OS   Bacillus anthracis.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,     Zamudio C,   Malone C,   Haselbeck R,  Ohlsen KL,   Zyskind JW;
PI   Wall D,     Trawick JD,  Carr GJ,    Yamamoto R,   Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA22620.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 46674; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
```

```
DE   Protein encoded by Prokaryotic essential gene #3163.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
OS   Bacillus anthracis.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,     Zamudio C,   Malone C,   Haselbeck R,  Ohlsen KL,   Zyskind JW;
PI   Wall D,     Trawick JD,  Carr GJ,    Yamamoto R,   Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA21506.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 45560; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note:The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 414 AA;

Query Match             13.2%;  Score 556.5;  DB 6;  Length 414;
Best Local Similarity   39.2%;  Pred. No. 1.2e-24;
Matches  145;  Conservative  54;  Mismatches 120;  Indels  51;  Gaps  14;

Qy     3  KSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENA 62
              |:||||  ||||| |||:|| ||||| |||| |   | :   :: || | :||||||||:|
```

```
CC  polypeptide or its fragment whose expression is inhibited by the
CC  antisense nucleic acid; (4) an antibody capable of specifically binding
CC  the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC  proliferation or the activity of a gene in an operon required for
CC  proliferation; (7) identifying a compound that influences the activity of
CC  the gene product or that has an activity against a biological pathway
CC  required for proliferation, or that inhibits cellular proliferation; (8)
CC  identifying a gene required for cellular proliferation or the biological
CC  pathway in which a proliferation-required gene or its gene product lies
CC  or a gene on which the test compound that inhibits proliferation of an
CC  organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC  compound's activity; (11) a culture comprising strains in which the gene
CC  product is overexpressed or underexpressed; (12) determining the extent
CC  to which each of the strains is present in a culture or collection of
CC  strains; or (13) identifying the target of a compound that inhibits the
CC  proliferation of an organism. The antisense nucleic acids are useful for
CC  identifying proteins or screening for homologous nucleic acids required
CC  for cellular proliferation to isolate candidate molecules for rational
CC  drug discovery programs, or for screening homologous nucleic acids
CC  required for proliferation in cells other than S. aureus, S. typhimurium,
CC  K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC  the target prokaryotic essential genes. Note: The sequence data for this
CC  patent did not form part of the printed specification, but was obtained
CC  in electronic format directly from WIPO at
CC  ftp.wipo.int/pub/published_pct_sequences
XX
SQ  Sequence 525 AA;

Query Match             13.2%;  Score 553.5;  DB 6;  Length 525;
Best Local Similarity   31.8%;  Pred. No. 2.5e-24;
Matches  164;  Conservative  71;  Mismatches  172;  Indels  109;  Gaps  15;

Qy           1 AGKSFPDVPAGHWAEGSINYLIVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE 60
                  | ||  |||| ||| |||||:| :||||:|| :|:||  |||| : :|||| :| ||||
Db          20 AAKKFSDVPT--WAQQSVDYLVGKKALDGKPDGTFSPSEAVDRGSAAKMLAVVLGLPVDP 77

Qy          61 NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV 120
                  |:||||||  ||| ||||||||| |||||||| :||||||| :|||| |||||| |||:
Db          78 KAKPSFKDAQNHWAAPYIAAVEKAGVINGDGTGKFNPSSQINRASMASMLVQAYSLDKKI 137

Qy         121 NGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQPIALTDKKYGK 180
                ||||  ||||| || || |||||||| ||||||||| || :| :||||||||||| | ||
Db         138 IGELPTQFKDLEPHWGKKQANILVALEISKGTGNGWNPEGTVTRAEAAQPIAMAEKKIKQ 197

Qy         181 KDNAQAY--------------VTDV-----KVSEPTKLTLTGTGLDKLSA------- 211
                 :|:|||              |||||     |||
Db         198 STSKRMYMTRNVITYHPSLSSGITTVHHKPQNGKVTEQ----RADGWVKMLTSKREKW 252

Qy         212 DDVTLEGDKAVAIEASTDGTSAVVTLGGKVAPNKDLTVKVKNQSFVTKFPYEVCKLAVBK 271
                 ||  |  |||  | ||  |  || |     |    |||||| :|||:|| ||| |||| |
Db         253 DTSNRKNRNDLMKEFSTYGT------ASHSSKVLGTYNAQTVTVMEKEWLIRI-- 300

Qy         272 LTFDDDRAGQAIAFKLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIFEGGEATSTTG 331
                 |:|:|||||  |  |     | :| | |   |:|   || ||  |   ||||:|:|  |
Db         301 ------RVGAGFQW---------VDKNQLNPVKQE-----NFLEGK-AIIDPGHGGIDSG 340

Qy         332 KLAVGIKQGDYKVEVQ-----VTKRGGLTVSNTGIITVK-NLDTPASAIKNVVFALDAD 384
                 ||||:|||||:|:|:|     ||||||| |  || | || |:|  |||||  | |  |
Db         341 NVGYYEKESDTVLDVSLRLKKIIXAKAPFTVMFTRTDNTRPGVNSTDSLKKRVEFAQEHN 400

Qy         385 NDGVVNYGSKLSGKDFALNSONL----------------------VVGEKASLNKLVATIAGED 426
                 |: |  |  || | :|| |:|                        :: :||:| :|:::|: ::|
Db         401 GDIFVSIHANGSAEKNGQGGTETLYQSARAKVTNPHVEDSKLLAQKIS-DRLVAALGTKD 459

Qy         427 KVVDPGSISIKSSNHGIISVVNNYITAEAAGEATLT 462
                 ||  |             || ||| |  :|| |||:
Db         460 RGVKHQDL------------YVTRENTLPAVLT 480
```

```
RESULT 6
ABU17384
ID   ABU17384 standard; protein; 379 AA.
XX
AC   ABU17384;
XX
DT   19-JUN-2003 (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #2911.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Bacillus anthracis.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-0081542.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-0072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,    Zamudio C,   Malone C,   Haselbeck R,   Ohlsen KL,   Zyskind JW;
PI   Wall D,    Trawick JD,  Carr GJ,    Yamamoto R,    Forsyth RA,   Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA21254.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 45308; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential gene. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 379 AA;

Query Match             10.9%;   Score 459;   DB 6;   Length 379;
Best Local Similarity   32.5%;   Pred. No. 5.8e-19;
Matches  124;  Conservative  49;  Mismatches  110;  Indels  98;  Gaps  13;
```

```
Qy    5  FPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQP  64
            |||| |   |  :| :|||||  ||  ||  ::||||||||:::| |||| ||::|
Db   24  FPDVPA--WADKSVTYLVDKQVLSGYPDGTFGSSDTLDRASAATIMTTALGIHIDLNAKP  81

Qy   65  SFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGEL  124
            |||: :|:|  |: ||||||| :| ||| :|  ||| ||| | :|||:|||  ::|
Db   82  SFKDSQNHWGTPYMPAAEKAGSIKVEGKGIFNPSGKVTRAAMATMLVNAYKLQNKNTSNG  141

Qy  125  VTTFEDLLDHWGEEKANILNLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYGKKDNA  184
            ||:|  | |||  |:|   :|||||| |:|||||| ::||||||:|:|: ||
Db  142  QSKFEFKGHWGGKIPNTLIGFEISVGTDNGWQPNKFITRAEAAQLTAKTD--------  192

Qy  185  QAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVTLGGKVAPN  244
                 :       |:  |   |:        ||  |:|
Db  193  ---MLQYSHSNPLE----------------NKTIIIDPGHGGED---------PG  219

Qy  245  KDLTVKVKNQSFVTKFVYEVKKLAVEKLT--------FDDDRAGQAIAFKLNDEKGNADV  296
            || |  |     :||| :||||
Db  220  KD-TKGLPXKKIVLDTSLRIQKL-LEKHTPFTVLNSXSDTRTGHGSKSSL-QERG----  272

Qy  297  EYLNLANHDVKFVANNLDGSPANIFEG-GEATST-------------------TGKLA  334
            :   :     :| :::|:        |    :||                   || |
Db  273  KFAKTKQGDILIMGH------ANAFNGNGKRTETYYXXSKSEKTNPHVEKKPGFPGKIQ  326

Qy  335  V----------GIKGQDYKV  344
            :           || :||
Db  327  TRLVDALQTRDRGVKHGDLHV  347
```

RESULT 7
ABU18670
ID   ABU18670 standard; protein; 408 AA.
XX
AC   ABU18670;
XX
DT   19-JUN-2003 (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #4197.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Bacillus anthracis.
XX
PN   WO20027183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-0007851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,
PI   Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA22540.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 46594; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 408 AA;

```
Query Match             8.8%;  Score 369;  DB 6;  Length 408;
Best Local Similarity   31.0%;  Pred. No. 1.2e-13;
Matches  148;  Conservative  57;  Mismatches 172;  Indels 100;  Gaps  21;

Qy    3  KSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENA  62
            | ||||   | |  ||||| :||| ||| ||   | ||  |||||||  :| ||||| |||
Db   22  KFFPDVQQGPGAEYSINYLAEKGAVTGNEKGMFEPGKEITRAEAATMMAKILNLPIDKNA  81

Qy   63  QPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNG  122
            :||: |:  ||  | :|||||||||||:| |||| ||| |||:||: |:|||||:|:||:
Db   82  KPSYADSQKHWATPIIAAVEKAGVVKGTG-NGFEPDGKIDRVSMASLLVEAYKLESKVNR  140

Qy  123  -ELVTTFEDLLDH-WGEEKANIL-INLGISVGTGGKWEPNKS-VSRAEAAQFIALTDKKY  178
             | :   | :    |  :|   :| :|                :
Db  141  FPQPINFKSFKNHKRGKEKRXYFKXSWVFFVGPGQGRGPGXENFXXKAEPVHFIKRPD--  197

Qy  179  GKKDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVTLG  238
                         |:|  :|           | :      ||
Db  198  ------------SLKVGNPLV----------EKVVIIDPGH--------G  217

Qy  239  GKVAPNKDLTVKVKNQSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEKGN-ADVE  297
            | :  |||:     : |||:        |:    ||       ::  | :  :|| :
Db  218  GFDPGNPGGQGVE-----ESKIVFDT-SLRLQKLL--EKNTPLKALLTREENGNPGSNK  267

Qy  298  YLNLANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLITVS  357
            :  ||:  |||  :|     |:| ::  |    :|           :||
Db  268  NESLANR-VKFGQEN-----NADIFVSIHANSSEKHDGHGFXNYYK----KSKRGEETQI  318

Qy  358  NTGIITVKNLDTPASAIKN-VVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLN  416
                :  :| :      |:| |:  |         |::|::|::  | :|  :::
Db  319  E------KDSEVLAKKIQNRVVEALHTDRKI-------KD---DHSLYVVNNNTVP  359

Qy  417  KLVATIAGEDKVVDPGSISIKSSNHGIISVVNNYITARAAGEATLTI---KVGDVTK  470
              | :  | || :         |:|:   |      :||||| : :     ||| :|
Db  360  AVLTELAFINDIDNGKLATESGRQ--------IAAEAVYAGILDYYEWKGFDVSK  407
```

RESULT 8
AAR80530
ID   AAR80530 standard; protein; 1252 AA.
XX
AC   AAR80530;

```
XX
DT   22-DEC-1995  (first entry)
XX
DE   B. sphaericus SLP.
XX
KW   Surface layer protein; SLP; fusion protein; vaccine; antigen;
KW   surface expression; epitope.
XX
OS   Bacillus sphaericus.
XX
FH   Key              Location/Qualifiers
FT   Peptide          1..30
FT                    /label= Sig_peptide
XX
PN   WO9519371-A2.
XX
PD   20-JUL-1995.
XX
PF   13-JAN-1995;  95WO-EP000147.
XX
PR   14-JAN-1994;  94GB-00000650.
XX
PA   (SOLV ) SOLVAY SA.
XX
PI   Debiaere RY, Desomer J, Dhaese P;
XX
DR   WPI; 1995-263827/34.
DR   N-PSDB; AAQ99430.
XX
PT   Host cell expressing surface layer protein fusion protein - used for host
PT   presentation of antigens and vaccine prodn.
XX
PS   Disclosure; Fig 6; 95pp; English.
XX
CC   A probe based on the N-terminal sequence of B. sphaericus P-1 (LMG P-
CC   13855) surface layer protein was used to screen an HindIII- generated
CC   library to isolate the slp gene. Promoter regions of the gene are used in
CC   genetic constructs providing surface expression of heterologous proteins
CC   in P-1 hosts
XX
SQ   Sequence 1252 AA;

Query Match             8.8%;  Score 369;  DB 2;  Length 1252;
Best Local Similarity   22.9%;  Pred. No. 5.4e-13;
Matches  221;  Conservative 131;  Mismatches 347;  Indels 264;  Gaps 46;

Qy   13 WAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQPSFKDAK-N 71
         :|::: |||::|| ||:  :    :|::||:::|: : |||| | :|:|:: |||:| ::
Db   42 YAKEAVQALVDQGVIQGDTNGNFNPLNTVTRAQAAEIFTKALEL-EANGDVNFKDVKAG 99

Qy   72 IWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASPASMLVSAYNLKDKVNGELVTTFEDL 131
         :|::: ||||: || :||:   | :||| ||:|  :::|| |: |::  || |:::| |
Db  100 AWYYNSIAAVVANGIFEGVSATEFAPNKSLTTRSEAAKILVEAFGLEGEAD--LSEFADA 156

Qy  132 --LDHWGEEKANILINLGISVGT-GGKWEPNKSVSRAE-AAQFIALTDKKYGKKDNAQAY 187
           :: ::  :: : |  :|:|   || ||:|:|:|:|: :||| | | ::: |::| :
Db  157 SQVKPWAKKYLEIAVANGIFEGTDANKLNPNNSITRQDFALVFKRTVDKVEGETPEEAAF 216

Qy  188 V-----TDVKVS---EPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSA----VV 235
         |     |::|::   ||:|:|| | || |:|| |:: :||:  |:: : ||:|      :
Db  217 VKAINNTTVEVTFEEEVTNVQALNFKIEGLEIKNASVKQTNKKVVVLTTEAQRADKEYVL 276

Qy  236 TLGGK-------VAPNKDITVKVKNQSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAF 285
         || |        |||:  ||||| |||| | |  ::|| ||:| |||| |||||||:
Db  277 TLDGETIGGFKGVAAVVPTK---VELVSSAVQGKLGQEVKVQAKVTVAEBQSKAGIPVTF 333

Qy  286 ------------KLNDEKGNADVEYLNLANHDVKFVA----NNLDGSPANIFEGG 324
                     :|: |:|||| ||| : :: ::|::     ::| ||||:|| |
Db  334 TVPGNNNDGVVPTLTGEALTNEEGIATYSYTRYKEGTDEVTAYATGDRSKFSLGYVFWGV 393

Qy  325 EATSTTGKLAVG--IKQG---DYKVEVQVTKRGGLTVS---NTGIITVKNL--DTPASAI 374
         :|:::||::|:    :|    |||: ||| ::| |:    ||| | ||:   |||:|
Db  394 DTILSVBEVTTGASVNNGANKTYKVTYKNPKTGKPEANKTFNVGFVENMNVTSDKVANAT 453

Qy  375 KNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGBKASLNKLVATIAGEDKVVDPGSI 434
         :|  | | : ::    :|  :|::: | :         ::|:: |||:||:||::|| :
Db  454 VNGVKALQLSN-----GTALDAAQITTDSK-----------GEATFTVSGTNAAVTPVVY 497

Qy  435 SIKSSNHGIISVVNNYITAB----------------------AAGEATLTIKV-----GDV---TKDV 472
         ::::||| ||:: ||:|                         ::|||:|||:|     ||:
Db  498 DLHSTNN--STSNKKYSASALQTTASKVTFAALQABYTIELTRADNAGEVAAIGATNGR 554

Qy  473 KFFKVTTDSR-------KLVSVKANPDKLQVVQNKTLPVTFVVTTDQYGDPFGANTAAIKEV 525
         : :|:| |::       ::|:|| :|::  :: :  :|:|  : |:: :|  | |  |:
Db  555 EYKVIVKDKAVNLAKNEIVNVAFNEDKDRVISTVT-NAKFVDTD------------- 597

Qy  526 LPKTGVVAEGG-----LDVVTTDSGS-----IGTKT----------IGVTGNDVGBGTV--- 564
         :|:|  | ||      || ||||          |:|          ||:|::| | ||
Db  598 -PDTAVYFTGDRAKQISVKTNDKGEATFVIGSDTVNDYATPIAWIDINTSDAKQGDLDEG 656

Qy  565 ----------HFQ-----NGNG-----ATLGSLYVNVTEGNVAFKNFELVSKVGQ---------- 599
                   ||:      ||||     ::|:| :|:|| |:|: |::|::::| ||
Db  657 EPKAVAPISYFQAPYLDGSAIKAYKKSDLNKAVTKFDGSETAVFAAELVNQSGKKVTGTS 716

Qy  600 --------YGQ-SPDTKLDLNV----STTVEYQLSKYTSDRVYSDPENLEGYEVESKNL 645
                 |||  |||:|||| :    |||| : ||||| :|  :|:|:|| |::||:|:
Db  717 IKKATYTIYNTGANDIKVDNQVISPNRSYTVTYEATLSSTGTVITPAKNLEVTSVDGKTT 776

Qy  646 AVADAKIVGNKVVVTG--KTPGVDIHLTKNGATAG-KATVEIVQETIAIKSVNFKPVQ 701
         ||   ||: |:|       ||| :: |||||||| :||| |:: |::| :::|||::
Db  777 AV------KVIATGIAVNTDGK-DYAFTAKEBATATFTATNEVPN--------SYTGVA 819

Qy  702 TE-NFVEKKINIGTVLELEKSNLDDIVRGINLTKETQHKVRVVKSGAEQGKL--YLDRNG 758
         |: || ::|| :::: :|  :| : :: |:|: |:|::||:|| ||:::|:|    |:|:
Db  820 TQFNTADSGSNSNIWFAGKN----------------------PVKYAGVSGKTYKYFGANG 859

Qy  759 DAVFNAGDVKLGDVTVSQTSDSALPNFKADLYDTLTTKYTDKG--TLVFKVLKDKDVIT 815
         : ||:|| ||:|||||||||||||||||||||||||||||||:  |:|||||||||:|:
Db  860 NEVFGEA------------------------------AWFALLTQYATEGQKVTISYNV--DGDTVT 894

Qy  816 SEI 818
         ::|
Db  895 FKV 897
```

```
RESULT 9
AAW22863
ID   AAW22863 standard; protein; 921 AA.
XX
AC   AAW22863;
XX
DT   17-OCT-2003  (revised)
DT   27-FEB-1998  (first entry)
XX
DE   Bacillus stearothermophilus S-layer protein sbs-B.
XX
KW   S-layer; sbs-B; vaccine; adjuvant; carrier; hybridisation assay;
KW   molecular spinning nozzle; molecular laser.
XX
OS   Geobacillus stearothermophilus.
XX
FH   Key              Location/Qualifiers
FT   Peptide          1..31
FT                    /label= sig_peptide
FT   Peptide          32..921
FT                    /label= mat_peptide
XX
PN   DE19603649-A1.
XX
PD   07-AUG-1997.
XX
PF   01-FEB-1996;  96DE-01003649.
XX
PR   01-FEB-1996;  96DE-01003649.
XX
PA   (LUBI/) LUBITZ W.
```

```
PA   (SLEY/) SLEYTR U.
XX
PI   Lubitz W, Sleytr U, Kuen B;
XX
DR   WPI; 1997-394558/37.
DR   N-PSDB; AAT75488.
XX
PT   Preparation of S-layer proteins by expressing sbs-A gene in Gram negative
PT   bacterium - or new sbs-B gene in any host, also new recombinant proteins
PT   containing heterologous inserts, e.g. epitope(s), useful as vaccines and
PT   adjuvants.
XX
PS   Claim 26; Page 19-23; 31pp; German.
XX
CC   The present sequence is the Bacillus stearothermophilus PV72 S-layer
CC   protein, sbs-3. S-layer structures can be used as vaccines or adjuvants,
CC   particularly when they include a bacterial ghost that may contain
CC   additional epitopes in its membrane. Other uses of recombinant sbs-B,
CC   depending on the nature of the inserted peptide, are as an universal
CC   carrier for biotinylated reactants for use in immunological or
CC   hybridisation assays (the insert is streptavidin), to induce immune
CC   responses (epitopes), as a reagent for removing cytokine or toxin from
CC   serum (antigenic epitopes), as a molecular spinning nozzle
CC   (polyhydroxybutyrate synthase) and as a molecular laser (luciferase).
CC   (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 921 AA;

Query Match             8.7%;  Score 367.5;  DB 2;  Length 921;
Best Local Similarity   24.9%;  Pred. No. 4.4e-13;
Matches  220;  Conservative 121;  Mismatches 335;  Indels 209;  Gaps     47;

Qy       4 SFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQ  63
             |::||||   | :    :|:::   :::  : :: : |  :|:||:: ::|||| :||
Db      33 SFTDVAPQY--KDAIDFLVSTGATKGKTETKFGVYDETTRLDAAVILARVLKDVDNAKD  90

Qy      64 PSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGE 123
             |||:||:|| :: ||::| ::||        :||  :::    |: ::: :::  |||
Db      91 AGFTDVPKD-RAKYVNALVEAGVLNGKAPGKFGAYDPLTRVEMAKIIANRYKLK---ADD 146

Qy     124 LVTTFEDLLDHWGE------EKANILINLGISVGTGGKWEPNKSVSRABAAQFIALTD 175
             | ::|  :::          :: ||::|| :: :  |  ::|        ::|||||
Db     147 VKLPFTDVNDTWAPYVKALYKYEVTKRLKHQQASVHT-------KNITLRDFAQFVY--- 196

Qy     176 KKYGKKDNAQAYVTDVKVSBPTKLLTGTGLDKLSADDVT-LEGDKAVAIEAST---DGT 231
             ::| :: :  |::|    :::   :|:    ||:: :||  ||||:|::|::    |||
Db     197 ----RAVNINAVPEIVEVTAVNSTVKVTFNTQIADVDFTNFAIDNGLTVTKATLSRDKK 252

Qy     232 SAVVTLGGKVAPNKDLTVK------VKN------QSFVTKFVYEVKKLAVEKLTFDDDRAGQA 282
             |:||||  |||| :|:|:      |::       ||| ||||||:| :|||: ||||::||
Db     253 SVEVVVNKPFTRNQEYTITATGIKNLKGETAKELTGKFVFWSVCQDAVTVALNNSSLKVGEE 312

Qy     283 IAFKLNDEKGNADVEYLNLANHDVKFVANNLDGSPANI--FEGGEATSTTGKLAVGIKQG 340
             || :::||:  |  |:| |:|  |  | ::|||| ::   :|| :|:::|||::|:||||
Db     313 SGLTVKDQDGK----------DVVGAKVELTSSNTNIVVVSSGEVSVSAAAKV-TAVKPG 360

Qy     341 DYKVEVQVTKRGGLTVSNTGIITVKNLDTPASAIKNVFALDADNDGVVNYGSKLSGKDF 400
             | ||||:::|    :|:::||::|| :::: |  :|::| :||   :|  |||:|||:
Db     361 TADVTAKVTLPDGVLTNTFKVTVTEVPV--QVONQGFTL-VDN------LSNA--VDN 405

Qy     401 ALNSQNLVVGEKASLNKLVATIAGEDKVV-------DPGS------ISIKSSNHGIIS 445
             | ||:|||::| ||:|:||||  | ||       ||||       ||:|:::|| ||
Db     406 --PQNTVAFNKA-EKVTSMFAGETKTVAMYDTKNGDPETKEVDFKDATVRSLNPIIAT 460

Qy     446 VVNN----YITAEA--AGEATLTIKVGDVTK-------FKVTTDSRKLVSVK 487
             ||:    |||:|:  ||:| |::|:|| ||       |:| |: ||:::|:
Db     461 AAINGSELLVTANAGQSGKASFEVTLKDNTKKRTFTVDVKKDPVLQDIKVDAITSVKLSDEA 520

Qy     488 ANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEYLPKTGVVAEGGLDVVTTDSGSI 547
             | |:::| :::| :  | | :  |::   : ||: ::: |  | | :| |:|:||:|||
Db     521 VGGGEVEGVNQKTIKVSAV--DQYG-----------KEI--KFGT--KGKVTVTTNTEGL 563

Qy     548 GTKTIGVTGNDVGESTVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQYGQSPDTK 607
```

```
CC   as bioreactors. (I) can be produced in carrier-bound form without loss of
CC   function. This sequence represents the Bacillus stearothermophilus S-
CC   layer protein sbsB which is used to illustrate the method of the
CC   invention. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 921 AA;

Query Match                8.7%;   Score 367.5; DB 3;  Length 921;
Best Local Similarity     24.9%;   Pred. No. 4.4e-13;
Matches 220;  Conservative 121;  Mismatches 335;  Indels 209;  Gaps  47;

Qy     4 SFPDVPAGHWABGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQ   63
         |||  || :    |:||  | ||| : |   :   |: :  ||| |      || || ||
Db    33 SFTDVAPQY--KDAIDFLVSTGATKGKTETKFGVVDEITRLDAAVILARVLKLDVDNAKD   90

Qy    64 PSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASPASMLVSAYNLKDKVNGE  123
         : : || :   : :| | | |  ||  |  |   |  || :       |  ||:  ||
Db    91 AGFTDVPKD-RAKYVNALVEAGVLNGKAPGKFGAYDPLTRVEMAKIIANRYKLK--ADD  146

Qy   124 LVTTFEDLLDHWGE-------EKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTD  175
         | :|:|| |      |      |:| || | : :|   |           | ||
Db   147 VKLPFTDVNDTWAPVVKALYKYEVTKRLKHQQASVHT------KNITLRDFAQFVY---  196

Qy   176 KKYGKKDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVT-LEGDKAVAIEAST--DGT  231
         ::     | :  |:|    :  || | |  |      :|       || |      ||
Db   197 ---RAVNINAVPEIVEVTAVNSTTVKVTFNTQIADVDFTNFAIDNGLTVTKATLSRDKK  252

Qy   232 SAVVTLGGKVAPNKDLTVK-----VKN-----QSFVTKFVYEVKKLLAVEKLTFDDDRAGQA  282
         |  | |||| |  || || :    || :   |:|  | | ||  |||
Db   253 SVEVVVNKPFTRNQEYTIITATGIKNLKGETAKELTGKFWNSVQDAVTVALNNSSLKVGEE  312

Qy   263 IAFKLNDEKGNADVEYLNLANHDVKFVANNLDGSPANI-FEGGEATSTTGKLAVGIKQG  340
         :  :       |  :  :  ::|      |      |: |    |:     |  ||
Db   313 SGLTVKDQDGK---------DVVGAKVELTSSNTNIVVVSSGEVSVSAAKV-TAVKPG   360

Qy   341 DYKVEVQVTKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDF  400
         : ||:      :|: |:: | ||    :     |:|          |
Db   361 TADVTAKVTLPDGVVLTNTFKVTVTEVPV--QVQNQGFTL-VDN------LSNA---   405

Qy   401 ALNSQNLVGEKASLNKLVATIAGEDKVV----------DPGS-----ISIKSSNHGIIS  445
         |:| ||  ::|   : | | |  ||:         :      | :::
Db   406 ---PQNTVAFNKA--EKVTSMFAGETKTVAMYDTKNGDPETKPVDFKDATVRSLNPIIAT  460

Qy   446 VVNN----YITAEA--AGEATLTIKVGDVTK------DVK------FKVTTDSRKLVSVK  487
         | ||    ||  || ||| | |:|| :|        |:|       || |||||:||||
Db   461 AAINGSELLVTANAGQSGKASFEVTLKDNTKRTFTVDVKKDPVLQDIKVDATSVKLSDEA  520

Qy   488 ANPDKLQVVQNKTLPVTFVVTTDQYGPDPFGANTAAIKEVLPKTGVVAEGGLDVVTTDSGSI  547
         | || |||| ||| ||:  |:|||||       |||:  |:|  || | | ||| | || :|
Db   521 VGGGEVBGVNQKTIKVSAV--DQYG--------KEI--KFGT--KGKVTVTTNTEGL  563

Qy   548 GTKTIGVTGNDVGEGTVHFQNGNGATLGSLYVNVTBGNVAFKNFELVSKVGYQYGGQSPDTK  607
         | | |  ::|          :     |:                              :||
Db   564 VIKNVN-----SDNTIDFDSGNSAT----------DQFVVVA-----------TK  592

Qy   608 LDLNVSTTVEYQLSKYTSDRVVSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKV  667
         | :|    || :  :| |:    |: ::|   | |:|  | |  :|  |||| || | |:
Db   593 -DKIVNGKVEVKFKNASD---TTPTSTKTITVNVNVNVK-ADATPVGLDIV----APSKI  643

Qy   668 DIHLTKNGATAGKATVEIVQ-ETIAIKSVN--------FKPVQTE--NFVEKKINIGTV  715
         |: : :|  ::  |:| | | ||    |:        || || |  ||| ||| | ||
Db   644 DVNAPN--TASTADVDFINFESVEIYTLDSNGRRQKKVTPTATTLVGTKKKKVN-GNV  699

Qy   716 LELEKSNLDDIVKGINLTKETCHKVRVKSGAE------QGKLYLDRNGDAVFNAGDVKL  769
         | | :| ||:|  ||  |:|     |:: | |      || ||  :|| |  | || |
Db   700 LQFKGNE------ELTLSTSSSTGNVDGTAEGMTKRIPGK-YI--NSASVPASATVAT  748

Qy   770 GDVTVS-QTSDGALPNFKADLYDTLTKYTDKGTLVFKVLKDKDV  813
         | |||:     :|      |  | : |  |  | ||| |:||::
Db   749 SPVTVKLNSSDNDL------TBELIFGVIDPTQLV---KDEDI  783

RESULT 11
```

```
AAW93253
ID   AAW93253 standard; protein; 920 AA.
XX
AC   AAW93253;
XX
DT   17-OCT-2003 (revised)
DT   24-AUG-1999 (first entry)
XX
DE   B stearothermophilus sbsB protein.
XX
KW   sbsB protein; S-layer protein; Gram-negative; prokaryotic host cell;
KW   integration; cytoplasmic membrane; secretion; periplasmic space; toxin;
KW   eukaryotic host cell; vaccine; adjuvant; immunogenic epitope; luciferase;
KW   immunostimulant; cytokine; polyhydroxybutyrate; PHB synthase; body fluid;
KW   molecular laser; universal carrier molecule; monomolecular layer.
XX
OS   Geobacillus stearothermophilus.
XX
FH   Key             Location/Qualifiers
FH
FT   Peptide         1..31
FT                   /label= signal_peptide
FT   Protein         32..920
FT                   /note= "sbsB"
XX
PN   DE19732829-A1.
XX
PN   04-FEB-1999.
XX
PD   30-JUL-1997;   97DE-01032829.
XX
PF   30-JUL-1997;   97DE-01032829.
PR
XX
PA   (LUBI/) LUBITZ W.
XX
PI   Lubitz W, Resch S;
XX
DR   WPI; 1999-122189/11.
DR   N-PSDB; AAX22751.
XX
PT   Producing S-layer proteins in Gram-negative bacteria or eukaryotes -
PT   integrated into membranes or organelles or secreted into periplasma or
PT   growth medium, and nucleic acid encoding S-layer proteins with peptide
PT   insertions, used in vaccines or for enzymatic reactions.
XX
PS   Disclosure; Page 23-25; 34pp; German.
XX
CC   This invention describes a method for the production of a S-layer protein
CC   (I) which comprises (a) preparing a Gram-negative prokaryotic host cell
CC   transformed with nucleic acid (II) encoding (I), linked to a signal
CC   sequence (SS) that encodes a protein which causes at least one of (i)
CC   integration of (I) into the external or cytoplasmic membranes and/or (ii)
CC   secretion of (I) into the periplasmic space or extracellular medium, (b)
CC   culturing the cell to express (I) and (c) optionally recovering (I) from
CC   the membranes, periplasmic space and/or extracellular medium.
CC   Alternatively, a eukaryotic cell is used as host and then the SS, which
CC   is optional, promotes integration into the cytoplasmic membrane or an
CC   organelle and/or secretion into the extracellular medium. (I), and
CC   derived structures, may include a wide variety of polypeptide inserts and
CC   are useful as (i) vaccines or adjuvants (with immunogenic epitopes or
CC   immunostimulants inserts such as cytokines) (ii) as reactors (inserts are
CC   enzymes, e.g. polyhydroxybutyrate (PHB) synthase for use as a 'molecular
CC   spinnerette' for production of PBH or luciferase for use as molecular
CC   laser (when combined with substrate and oxygen) and (iii) as universal
CC   carrier molecule (streptavidin is inserted) for use in hybridisation and
CC   immuno assays, or for selective elimination of cytokines, toxins etc.
CC   from body fluids (inserts are specific binding epitopes). In this system,
CC   heterologous (I) do not form inclusion bodies but rather monomolecular
CC   layers, and in eukaryotic cells they undergo glycosylation. This sequence
CC   represents the Bacillus stearothermophilus sbsB protein which is used to
CC   illustrate the method of the invention. (Updated on 17-OCT-2003 to
CC   standardise OS field)
XX
SQ   Sequence 920 AA;
```

```
RESULT 12
AAR41731
ID   AAR41731 standard; protein; 1338 AA.
XX
AC   AAR41731;
XX
DT   25-MAR-2003 (revised)
DT   26-APR-1994 (first entry)
XX
DE   High molecular weight protein 3 (HMW3).
XX
KW   HMW; high molecular weight protein; virus; vaccine; influenza; epitope;
KW   immunity; haemophilus influenzae.
XX
OS   Haemophilus influenzae.
XX
PN   WO9319090-A1.
XX
PD   30-SEP-1993.
XX
PF   16-MAR-1993; 93WO-US002166.
XX
PR   16-MAR-1992; 92GB-00005704.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1993-320683/40.
DR   N-PSDB; AAQ49510.
XX
PT   High molecular weight surface proteins - of non-typeable haemophilus
PT   which exhibit immunogenic properties.
XX
PS   Claim 5; Fig 10; 100pp; English.
XX
CC   The isolation and purification of the high molecular weight protein
CC   enables the identification of the major protective epitopes of the
CC   protein by conventional epitope mapping. These epitopes can then be
CC   synthesised using standard techniques and incorporated into fully
CC   synthetic or recombinant vaccines. (Updated on 25-MAR-2003 to correct PN
CC   field.)
XX
SQ   Sequence 1338 AA;

Query Match       6.1%; Score 257.5; DB 2; Length 1338;
Best Local Similarity  22.0%; Pred. No. 2.1e-06;
Matches 183; Conservative 104; Mismatches 328; Indels 215; Gaps 38;

Qy   16 GSI------NYLVDKGA--------------ITGKPDGTYGPTESIDRASAAVIFTKILN--L 56
Db  631 GSIINIAGNLTVSKGANLQAITNYTFNVAGSFDNNGASNISIARGGAK--FKDINNTSSL 688

Qy   57 PVDENAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNL 116
Db  689 NITTNSDTTYR----------------TIIKGNISNKSGDLNI----IDKKSDAEIQIGG-NI 730

Qy  117 KDKVNGELVTTPEDLLDHWGEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDK 176
Db  731 SQK-EGNLTIS----------SDKVNITNQIITKAGVEG----------- 758

Qy  177 KYGKKDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVT 236
Db  759 --GRSDSSEA----------ENANLTIQTKEL-KL-AGDLNISGFNKAEITAKNGSDLTIGN 806

Qy  237 LGGKVAPNKDLTV-KVKNQSFVT----KFVVEVKKLAVEKLTFDDDRAGQAIAFKLNDE 290
Db  807 ASGGNADAKKKVTFDKVKDSKISTDGHNVTLNSEVKTSNGSNAGNDNSTGLTISAK---- 862

Qy  291 KGNADVEYLNLANHDVKFVANNLDGSPANIF-EGGEATSTTGKLAVGIKQGDYKVEVQV 348
Db  863 ----DV---TVNNNVTSHKTINISAAAGNVTTKEGTTINAITTGSVEVTAQNG-------- 907
```

```
Query Match       8.7%; Score 367; DB 2; Length 920;
Best Local Similarity  23.3%; Pred. No. 4.7e-13;
Matches 227; Conservative 138; Mismatches 369; Indels 240; Gaps 48;

Qy    4 SFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQ 63
Db   33 SFTDVAPQY--KDAIDFIVSTGATKGKTETKFGVYDEITRLDAAVILARVLKLDVDNAKD 90

Qy   64 PSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGE 123
Db   91 AGFTDVPKD-RAKYVNALVEAGVLNGKAPGKFGAYDPLTRVEMAKIIANRYKLK--ADD 146

Qy  124 LVTTFEDLLDHWGEEKANILINLGISVG-TGGKWEPNKSVSRAEAAQFIALTDKKYGKKD 182
Db  147 VKLPFTVDNDTWA-PYVKALYKYEVTKGKTPTSFGAYQNITRGDFAQPFVY------RAV 198

Qy  183 NAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVT-LEGDKAVAIEAST--DGTSAVVTLG 238
Db  199 NINAVPEIVEVTAVNSTTVKVTFNTQIADVDFTNFAIDNGLTVTKATLSRDKKSVEVVVN 258

Qy  239 GKVAPNKDLTVK---VKN------QSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAFKLND 289
Db  259 KPFTRNQEYTITATGIKNLKGETAKELTGKFVWSVQDAVTVALNNSSLKVGEESGLTVKD 318

Qy  290 EKGNADVEYLNLANHDVKFVANNLDGSPANI--FEGGEATSTTGKLAVGIKQGDYKVEVQ 347
Db  319 QDGK-----------DVVGAKVELTSSNTNIVVVSSGEVSVSAAKV-TAVKPGTADVTAK 366

Qy  348 VTKRGGLITVSNTGIITVKNLDTPASAIKNVFALDADNDGVVNYGSKLSGKDFALNSQNL 407
Db  367 VTLPDGVVLTNTFKVTVTEVPV--QVQNQGFTL-VDN------LSNA-----PQNT 408

Qy  408 VVGEKASLNKLVATIAGEDKVV-----------DPGS------ISIKSSNHGIISVVNN-- 449
Db  409 VAFNKA--EKVTSMFAGETKTVAMYDTKNGDFETKPVDFKDATVRSLNPIIATAAINGSE 466

Qy  450 -YITAEA--AGEATLTIKVGDVTK-----DVK------FKVTTDSRKLVSVKANPDKLQ 494
Db  467 LLVTANAGQSGKASFEVTFKDNTKRTFTVDVKKDPVLQDIKVDATSVKLSDEAVGGGEVE 526

Qy  495 VVQNKTLPVTFVVTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVVTTDSGSIGTKTIGV 554
Db  527 GVNQKTIKVSAV--DQYG---------KEI--KFGT---KGKVTVTTNTEGLVIKNVN- 568

Qy  555 TGNDVGEGTVHFQNGNGATLGSLYV----NVTEGNVAFKNFELVS----------- 595
Db  569 -----SDNTIDFDSGNSATDQFVVATDKIVNGKVEVKYFKNASDTTPTSTKTITVNVV 623

Qy  596 ------KVGQYGQGSPDTKLDLNVSTTVEYQLSKYTSDRVYSGDPENLEGYEVBS-----K 643
Db  624 NVKADATPVGLDIVAP-SEIDVNAPNTAS------TADVFINFESVEIYTLDSNGNRLK 676

Qy  644 NLAVADAKIVGNK--VVVTGK--------------TPGKVDIHLTKNGAT----AGK 680
Db  677 KVTPTATTLVGTNDYVEVNGNVLQFKGNDELTLLTSSSTVNVDVTADGITKRIPVKYINS 736

Qy  681 ATVEIVQETAIKSVNFKPVQTEN---FVEKKINIGTVLELEKSNLDDIVKGINLTKETQ 737
Db  737 ASVP-ASATVATSPVTVKLNSSDNDLTFEELIFGVIDPTQLVKD--EDINEFIAVSKAAK 793

Qy  738 H.---------------KVRVVKSGAE-------QGKLYLDRN--------GDAVFN-- 763
Db  794 NDGYLYNKPLVTVKDASGKVIPTGANVYGLNHDATNGNIWFDEEQAGLAKKFSDVHFDVD 853

Qy  764 --AGDVKLGDVTVSQTSDSALPNFKADLYDTLTTKYTDKG-----TLVFK--VLKDKDV 813
Db  854 FSLANVVKTGSGTVSSS-------------PSLSDAI--QLTNSGDAVSFTLVIKSIYVKGADK 902

Qy  814 ITSEIGSQAVHVNV 827
Db  903 DDNNLLAAPVSVNV 916
```

```
XX
PS   Claim 1; Page 93-97; 183pp; English.
XX
CC   This protein comprises the high molecular weight surface protein HMW3
CC   (125 kDa) of non-typeable Haemophilus influenzae strain 5 that has the
CC   immunological ability to protect against disease caused by a non-typeable
CC   Haemophilus strain and is characterised by at least one surface-exposed B
CC   -cell epitope that is recognised by monoclonal antibody AD6. The HMW3
CC   amino acid sequence was deduced from an isolated antibody hmw3 (see
CC   AAT90992). HMW1 (see AAW30293), HMW2 (see AAW30294) and HMW4 (see
CC   AAW30292) have also been identified. A conjugate comprising HMW3 linked
CC   to an antigen, hapten or polysaccharide, and a synthetic peptide of 6-150
CC   amino acids corresponding to at least protective epitope of HMW3 are also
CC   claimed. HMW proteins, conjugates and peptides can be used in vaccines,
CC   as immunogens for preparation of antibodies and as antigens for detection
CC   of these antibodies. (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1598 AA;

Query Match             6.1%;  Score 257.5;  DB 2;  Length 1598;
Best Local Similarity  22.0%;  Pred. No. 2.6e-06;
Matches 183;  Conservative 104;  Mismatches 328;  Indels 215;  Gaps 38;

Qy    16 GSI-----NYLVDKGA-----------------ITGKPDGTYGPTESIDRASAAVIFTKILN--L    56
         |||                              |:|||
Db   891 GSIINIAGNLTVSKGANLQAITNVTFNVAGSFDNNGASNISIARGGAK--FKDINNTNSL       948

Qy    57 PVDENAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNL      116
Db   949 NITTNSDTTYR------------TIIKGNISNKSGDLNI-----IDKKSDABIQIGG-NI      990

Qy   117 KDKVNGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDK      176
Db   991 SQK-EGNLTIS------------------SDKVNITNQITIKAGVEG-----------       1018

Qy   177 KYGKKDNAQAYVTDVKVSEPTKLTLTGLDKLSADDVTLEGDKAVA1EASTGTSAVVT        236
Db  1019 --GRSDSSEA------------ENANLIIQTKEL-KL-AGDLNISGFNKAEITAKNGSDLTIGN   1066

Qy   237 LGGKVAPNKDLTV-KVKNQSFVT-----KFVYBVKKLAVEKLTFDDDRAGOAIAFKLNDE      290
         |||                       |:|||
Db  1067 ASGNADAKKVTFDKVKDSKISTDGHNVTLNSBVKTSNGSSNAGNDNSTGLTISAK----      1122

Qy   291 KGNADVEYLNLANHDVKFVANNLDGSPANIP--EGGBATSTTGKLAVGIKQGDYKVEVQV      348
Db  1123 ----DV--TVNNNVTSHKTINISAAAGNVTTKEGTTINATTGSVEVTAQNG----          1167

Qy   349 TKRGGLTVSNTGIITVKNLDTPASAIKNVVPALDADNDGVVNYGSKLSGKDFALNSQNLV      408
         |:||  |: |:  :  |:    |   |:           :     |:
Db  1168 TIKGNITSQNVTVTATENLVTTENAVINA------TSGTVNISTKTIGDIKGGIESTSGN      1220

Qy   409 VGEKASLNKL-VATIAGEDKVVDPGSISIKSSNHGIISVVNNYITABAAGEATLTIKVGD      467
         |:||:|::| |:||||||:
Db  1221 VNITASGNTLKVSNITGQDVTV-----TADAGALTTTAGSTISATTGNANITTKTGD       1272

Qy   468 VTKDVK------------FKVTTDSRKLVSVKANPDKLQVVQNKTLPVT--FVTTDQY       511
         |:|:||
Db  1273 INGKVESSSGSVTLVATGATLAVGNISGNTVTITADSGKLTSTVGSTINGTNSVTTSSQS    1332

Qy   512 GDPFGANTAIKEVLPKTG------------VVAEGGLDVVTTDSGSIGTKTIGVTGNDVGEG    562
         |||:|:
Db  1333 GDIEGTISGNTVNVTASTGDLTIGNSAKVEAKNGAATLTAESGKLTTQ---TGSSI---    1385

Qy   563 TVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQYGQSPDTKLDLNVST-TVEYQLS    621
Db  1386 ---TSSNGQTTLTAKDSSIAGNINAAANVTL-NTTGLTTTGDSKINATSGTLTINAKDA    1440

Qy   622 KY----TSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVDIH--LTKN    674
Db  1441 KLDGAASGDRTVVNATNASG------SGN--VTAKTSSSVNITGDLNTIN            1482

Qy   675 G----ATAGKATVEIVQEFIAIKSVNFKPVQTENFVBKKINIGTVLELEKSNLDDIVK-G    729
```

```
Qy   349 TKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLV      408
         |:||  |: |:  :  |:    |   |:           :     |:
Db   908 TIKGNITSQNVTVTATENLVTTENAVINA------TSGTVNISTKTIGDIKGGIESTSGN       960

Qy   409 VGEKASLNKL-VATIAGEDKVVDPGSISIKSSNHGIISVVNNYITABAAGEATLTIKVGD      467
         |:||:|::| |:||||||:
Db   961 VNITASGNTLKVSNITGQDVTV-----TADAGALTTTAGSTISATTGNANITTKTGD       1012

Qy   468 VTKDVK------------FKVTTDSRKLVSVKANPDKLQVVQNKTLPVT--FVTTDQY       511
         |:|:||
Db  1013 INGKVESSSGSVTLVATGATLAVGNISGNTVTITADSGKLTSTVGSTINGTNSVTTSSQS    1072

Qy   512 GDPFGANTAIKEVLPKTG------------VVAEGGLDVVTTDSGSIGTKTIGVTGNDVGEG    562
         |||:|:
Db  1073 GDIEGTISGNTVNVTASTGDLTIGNSAKVEAKNGAATLTAESGKLTTQ---TGSSI---    1125

Qy   563 TVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQYGQSPDTKLDLNVST-TVEYQLS    621
Db  1126 ---TSSNGQTTLTAKDSSIAGNINAAANVTL-NTTGLTTTGDSKINATSGTLTINAKDA    1180

Qy   622 KY----TSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVDIH--LTKN    674
Db  1181 KLDGAASGDRTVVNATNASG------SGN--VTAKTSSSVNITGDLNTIN            1222

Qy   675 G----ATAGKATVEIVQEFIAIKSVNFKPVQTENFVBKKINIGTVLELEKSNLDDIVK-G    729
Db  1223 GLNIISENGRNTVRLRGKEIDVKYIQPGVASVEEVIEAKRVLEKVKDLSDEERETLAKLG    1282

Qy   730 IN--------LTKETQHKVRVVKSG--ABQGKLYLDRNGDAVFNAGD               766
         ||                                |:|:
Db  1283 VSAVRFVEPNNAITVNTQNEFTTKFPSSQVTISEGK------ACFSSGN            1324
```

```
RESULT 13
AAW30291
ID   AAW30291 standard; protein; 1598 AA.
XX
AC   AAW30291;
XX
DT   17-OCT-2003 (revised)
DT   14-APR-1998 (first entry)
XX
DE   Non-typeable Haemophilus high mol.wt. surface protein HMW3.
XX
KW   Non-typeable Haemophilus; high molecular weight surface protein; HMW3;
KW   immunogen; vaccine; otitis media.
OS   Haemophilus influenzae; strain 5.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 113
FT                   /note= "encoded by GTG"
FT   Misc-difference 864
FT                   /note= "encoded by TGT"
XX
PN   WO9736914-A1.
XX
PD   09-OCT-1997.
XX
PF   01-APR-1997; 97WO-US004707.
XX
PR   01-APR-1996; 96US-00617697.
XX
PA   (BARE/) BARENKAMP S J.
PI   Barenkamp SJ;
XX
DR   WPI; 1997-503038/46.
DR   N-PSDB; AAT90992.
XX
PT   High molecular weight proteins of non-typeable Haemophilus influenzae -
PT   useful for vaccine production.
```

```
Db   1483 GLNIISENGRNTVRLRGKEIDVKYIQPGVASVEEVIEAKRVLEKVKDLSEERETLAKLG 1542
                                                                        
Qy    730 IN--------LTKETQHKVRVVKSG--AEQGKLYLDRNGDAVFNAGD 766
          ::  |||||:                    :||:::
Db   1543 VSAVRFVEPNNAITVNTQNEFTTKPSSQVTISEGK------ACPSSGN 1584
```

RESULT 14
AAR77395
ID   AAR77395 standard; protein; 941 AA.
XX
AC   AAR77395;
XX
DT   10-APR-1996 (first entry)
XX
DE   Full length Bacillus sp. alkaline cellulase.
XX
KW   Alkaline cellulase; Bacillus; E.coli; pBR322; expression vector; stable;
KW   B.subtilis; specific activity; surface active agent; chelating agent.
XX
OS   Bacillus sp.
XX
PN   JP07203960-A.
XX
PD   08-AUG-1995.
XX
PF   24-JAN-1994; 94JP-00005714.
XX
PR   24-JAN-1994; 94JP-00005714.
XX
PA   (KAOS) KAO CORP.
XX
DR   WPI; 1995-307162/40.
DR   N-PSDB; AAQ94350.
XX
PT   Alkaline cellulase and related DNA, vectors and transformed microbes -
PT   useful for prodn. of a highly specific enzyme in alkaline environments.
XX
PS   Disclosure; Page 10-14; 17pp; Japanese.
XX
CC   The amino acid sequence of the full length alkaline cellulase. The gene
CC   encoding the protein was isolated from Bacillus sp. KSM-365 chromosomal
CC   DNA and inserted into the E.coli vector pBR322. The gene was subsequently
CC   used to produce expression vectors contg. a 1071 fragment of the gene
CC   covering amino acid residues 228-584 (AAQ94349). The expression vectors
CC   contg. the fragment were transformed into B.subtilis for production of
CC   the protein. The novel alkaline cellulase has high specific activity in
CC   an alkaline environment and is stable over wide ranges of temp. and pH,
CC   and has full activity in the presence of surface active agents and
CC   chelating agents
XX
SQ   Sequence 941 AA;

Query Match   6.0%;   Score 252.5; DB 2; Length 941;
Best Local Similarity 21.3%;   Pred. No. 2.6e-06;
Matches 193;   Conservative 109;   Mismatches 291;   Indels 311;   Gaps 42;

Qy   5 FPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQP 64
         |||||||||| :   :||| ::|:||| |:||:|:||| |:| :|:|
Db   41 FSDVKKTSWSFPYIKDLYEQEVITGTSATTFSPTDSVTRAQFTVMLTRGLGEASSKDYP 100

Qy   65 SFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNL-------- 116
         :|||||||:||::|:||:| ||: :|||:||||:||| |:| :|
Db   101 -FKDRKN-WAYKEIQAAYEAGIVTGKTNGEFAPNENITREQMAAMAVRAVEYLENELSLP 158

Qy   117 --KDKVNGELVTTF-EDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRABAAQFIA 172
           :|| :| |:::| ||:|:| ||:||:||:|:|:: |:| ::|:|
Db   159 EEQREYNDSSISTFAQDAV-----QKAYVLEIM--EGNTDGYFQPKRNSTREQSAKVIS 211

Qy   173 LTDKKYGKKDNAQAYVTDVKVSEPTK------LTLTGTGLDKLSADDVTLEGDKAVAIE- 225
         |:|||:|:|| :|:|:|| :| |:||||| :||||| |||| |
Db   212 TLLWKVASHD--YLYHTEA-VKSPSEARGALQVLELNG--------QLTLAGEDGTPVQL 259
```

```
Db    226 --ASTDGT---SAVVTLGKGVAPNKD--------LTVKVNQSFVTK-----FVYEVKKLA 268
                                                    :: |::        ::|
Qy    260 RGMSTHGLQWFGEIVNENAFVALSNDWGSNMIRLAMYIGENGYATNPEVKDLVYEGIELA 319
                                                    |: :|||  : |||  :||:
Db    269 VEKLTF--------DDDRAG---QAIAFKLNDEKGNADVEYL--NLANHDVK 307
          |||:|        ||||||    |::      :|  | :|    :||:| |
Qy    320 FEHDMYIVDWHVHAPGDRADVYSGAYDFFEEIADHYKDHPKN--HYIIWELANEP-- 374
          |  |    |||   |||||||    |::::|:|||||:|:|:    :|  |||:||
Db    308 FVANNLDGSP--ANIFEGEATSTTGKLAVGI--KQGDYKVEV---QVTKRGGLTVSNTG 360
          |  ||| :|    |:||| |:::|||  |    :|||||:|    ::||:|| ||:  
Qy    375 --SPNNNGGPGLTNDEKGWBAVKEYAEPIVEMLREKGDNMILVGNPNWSQRPDLSADNP- 431
            :|:| || |::|:|:| |: || || | ::| :: || ::|||  |:|| ::|
Db    361 IITVKNL-----------------------DTPAS----AIKNVVFALD------- 382
          ||||:||                        |:| |     :|:|:|:||
Qy    432 -IDAENIMYSVHFYTGSHGASHIGYPEGTPSSERSNVMANVRYALDNGVAVFATEWGTSQ 490
           :| |:| ||| |||| |: |: |:|||:| |::|:||  |||:| |: | | |  ||
Db    383 ADNDG--------VVNYGSK--LSGKDFALNSQNLVVG--EKASLNKLVATIAGEDKVV 429
          |:|         |:|| |:  |:|:||:||:|||:|    :|||||:||:| ||| |:
Qy    491 ANGDGGPYFDEADVWLNFLNKHNISWANWSLTNKNEISGAFTPFLGRTDAT------DL 544
          |: || |:|||||| |::||:| ||| |:|| ||:||:| || || ||||       |
Db    430 DPGSISIKSSNHGIISVVNNYITAEAAGEATLTIKVGDVTKDVK-------------FKVT 477
          ||| |:|:|||||:|||:|||::|:| ||  ||:| ||:||||             |:||
Qy    545 DPGANQVWAPEE--LSLSGEYVRARIKGIEYTPI---DRTKFTKLVWDFNDGTTQGFQVN 599
          |||| | | |    |:||| | || || |||     ||::|||:|||:|||:||||||||
Db    478 TDS--RKLVSVKANPDKLQV----VQN---------- 498
          |:|   |:||: |||:|:|     |:|
Qy    600 GDSPNKESITLSNNNDALQIEGLNVSNDISEGNYWDNVRLSADGWSENVDILGATELTID 659
          ||||||:|||  |:|:|| :||||||:|:||||||:|::|||||||:||||||||||||
Db    499 ----------------KTLPVTFVVTDQYGDPFGANTAIKEVLPKTGVV 532
                          |::||||||||||||||:||||:|:|||||||||
Qy    660 VIVEPTTVSIAAIPQGPAAGWANPTRAIKVTEDDFESFGDGYKALVTTSEDSPSLETI 719
          |:|||||||:||:|| |||| |:||:||||||||::|||||||||:||:|||:|:||
Db    533 AEGGLD----------VVTTDSGIGTKTIGVTGNDVGEGTVHFQNGNGATLGSLY----- 578
          |:|||            ||||||:||:|:||||:||||||||:|:|||||||||
Qy    720 ATSPEDNTMSNIILFVGTEDADVISLDNITVSGTEIBIEVIHDEKGT-ATLPSTPEDGTR 778
          |:||:||||:|||:||||:||||:|||||||:|||:|:||||::||  |||||||||||
Db    579 --------------VNVTEGN------------VAFKNFELVSKVGQ 599
                        :|||||              |||||:|||||||||
Qy    779 QGWDWHTESGVKTALTIEEANGSNALSWEYAYPEVKPSDGWATAPRLDFWKDELVRGTSD 838
          ||||||||||||||||||::| |||||:|||||:||||||| ||||||||:||||||||
Db    600 Y----------------GQSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEG 637
          |                ||||||||:|||:|||||||||:|||||||:||:|||
Qy    839 YISPDFYIDAVRASBGAISINAVFQPPANGVWQBVPTTFEIDLTELDSATVTSD--ELYH 896
          |||||||:|||:|:||:||||||:|||| |||:||:|||:|||:|||:||:|||   |||
Db    638 YEVE 641
          ||||
Qy    897 YEVK 900
          ||||
```

RESULT 15
AAR07478
ID   AAR07478 standard; protein; 941 AA.
XX
AC   AAR07478;
XX
DT   24-OCT-2003   (revised)
DT   27-AUG-2003   (revised)
DT   06-FEB-1991   (first entry)
XX
DE   Cellulase.
XX
KW   Enzyme.
XX
OS   Bacillus sp; KSM-634.
XX
FH   Key   Location/Qualifiers
FT   Protein   1..584
FT   /note="claim 2"
XX
PN   JP01281090-A.
XX
PD   13-NOV-1989.
XX

```
PF   02-MAY-1988;   88JP-00109545.
XX
PR   02-MAY-1988;   88JP-00109545.
XX
PA   (KAOS ) KAO CORP.
XX
DR   WPI; 1990-330487/44.
DR   N-PSDB; AAQ06320.
XX
PT   DNA segment contg. cellulase gene - originated from alkalophilic Bacillus
PT   and codes cellulase.
XX
PS   Claim 1; Fig 1; 21pp; Japanese.
XX
CC   DNA encoding the sequence can be introduced into E.coli to give a
CC   recombinant strain for prodn. of cellulase protein. (Updated on 27-AUG-
CC   2003 to correct OS field.) (Updated on 24-OCT-2003 to standardise OS
CC   field)
XX
SQ   Sequence 941 AA;

Query Match             6.0%;   Score 251.5;  DB 2;   Length 941;
Best Local Similarity  21.3%;   Pred. No. 2.9e-06;
Matches  193;  Conservative 108;  Mismatches 292;  Indels 311;  Gaps  42;

Qy      5 FPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQP  64
          :|:  | |: ||| |  |: |   : |: |     |: |  |||| |  :||    |  |
Db     41 FSDVKKTSWSPPYIKDLYEQEVITGTSATTFSPTDSVTRAQFTVMLTRGLGLEASSKDYP 100

Qy     65 SFKDAQNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNL--------- 116
          ||||| ||||||  ||:|:  ||:|||:  ||: ||||||||  ||||:||
Db    101 -FKDRKN-WAYKEIQAAYEAGIVTGKTNGEFAPNENITREQMAAMAVRAYEYLENELSLP 158

Qy    117 ---KDKVNGELVTTF-EDLLDHWGEEKANLINLGISVGTGGKWEPNKSVSRAEAAQFIA 172
             |||  | |||:  |||| ||| ||||||||||||||| || |: |:  |  : |
Db    159 EEQREYNDSSSISTFAQDAV-----QKAYVLELM-EGNTDGYFQPKRNSTREQSAKVIS 211

Qy    173 LTDKKYGKKDNAQAYVTDVKVSEPTK------LTLTGTGLDKLSADDVTLEGDKAVAIE- 225
          |  :| |:| ||  |:   :  | :          | || |:|||| |:|| |: |  |:
Db    212 TLLWKVASHD--YLYHTEA-VKSPBAGALQLVELNG---------QLTLAGEDGTPVQL 259

Qy    226 --ASTDGT--SAVVTLGGKVAPNKD---------LTVKVKNQSFVTK-----FVYEVKKLA 268
            ||:||    :| ||||| | ||          | ||:|:|| |         |:|| ||
Db    260 RGMSTHGLQWFGEIVNENAFVALSNDWGSNMIRLAMYIGENGYATNPEVKDLVVEGIELA 319

Qy    269 VEKLTF-----------DDDRAG---------QAIAFKLNDEKGNADVEYL--NLANHDVK 307
          ||| :            ||| |          :|  | ||:|:|   |:|   |:|||:||
Db    320 FEHDMYVIVDWEVHAPGDPRADVYSGAYDFFEIADHYKDHPKN--HYIIWELANEP-- 374

Qy    308 FVANNLDGSP--ANIFEGGEATSTTGKLAVGI--KQGDYKVEV--QVTKRGGLTVSNTG 360
          || |  ||:    ||   | |:|| ||| |    |:|||:|    | |:||| || |||
Db    375 --SPNNNGGPGLTNDEKGWEAVKEYAEPIVEMLREKGDNMILVGNPNWSQRPDLSADNP- 431

Qy    361 IITVKNL------------DTPAS----AIKNVVFALD--------------------- 382
          | ||||             | |||    ||||||| ||
Db    432 -IDAENIMYSVHFYTGSHGASHIGYPEGTPSSERSNVMANVVLLLDNGVAVFATEWGTSQ 490

Qy    383 ADNDG-------VVNYGSK--LSGKDFALNSQNLVVG--EKASLNKLVATIAGEDKVV 429
             | |       ||| |:   ||| | ||:|||||    ||||:|| || ||||:||
Db    491 ANGDGGPYFDEADVWLNFLNKHNISWANWSLTNKNEISGAFTPFELGRTDAT------DL 544

Qy    430 DPGSISIKSSNHGIISVVNNYITAEAAGEATLTIKVGDVTKDVK--------FKVT 477
          ||| :|  |     |: :   |    || : |    | ||:      |:|
Db    545 DPGANQVWAPEE--LSLSGEYVRARIKGIEYTPI---DRTKFTKLVWDFNDGTTQGFQVN 599

Qy    478 TDS--RKLVSVKANPDKLQV----VQN------------------------------ 498
           |:   :|   | || |:       |:
Db    600 GDSPNKESITLSNNNDALQIEGLNVSNDISEGNYWDNVRLSADGWSENVDILGATELTID 659

Qy    499 ----------------KTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVV 532
                            | | | | |: ::  | |  :|| || : |  |||
Db    660 VIVEEPTTVSIAAIPQGPAAGWANPTRAIKVTEDDFESFGDGYKALVTITSEDSPSLETI 719
```

```
Qy    533 AEGLD--------VVTTDSGSIGTKTIGVTGNDVGEGTVHFQNGNGATLGSLY----- 578
              |:|| |        ||  :|||| ||| |  |||| | : |  : |     |||||||
Db    720 ATSPEDNTMSNIILFVGTEDADVISLDNITVSGTEIEIEVIHDEKGT-ATLPSTFPEDGTR 778

Qy    579 ----------------VNVTEGN------------------VAFKNFELVSKVGQ 599
                              ::|:|:||
Db    779 QGWDWHTESGVKTALTIEEANGSNALSWEYAYPEVKPSDGWATAPRLDFWKDELVRGTSD 838

Qy    600 Y------------------GQSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEG 637
          |                  ||| :  |  |  |  |: |   ::   |  ||||: |:
Db    839 YISFDFYIDAVRASEGAISINAVFQPPANGYWQEVPTTFEIDLTELDSATVTSD--ELYH 896

Qy    638 YEVE 641
          ||||
Db    897 YEVK 900
```

Search completed: April 7, 2004, 17:26:34
Job time : 68 secs

```
                 Copyright   GenCore version 5.1.6
                 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   April 7, 2004, 17:23:54 ; Search time 46 Seconds
                                     (without alignments)
                                     4755.947 Million cell updates/sec

Title:         US-09-844-281-1
Perfect score: 4202
Sequence:      1 AGKSFPDVPAGHWAEGSINY.........ITSEIGSQAVHVNVLNNPNL 833

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:      1071772 seqs, 262633353 residues

Total number of hits satisfying chosen parameters:    1071772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  .Published Applications AA:*
        1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
        2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
        3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
        4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
        5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
        6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
        7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
        8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
        9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
       10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
       11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
       12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
       13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
       14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
       15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
       16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
       17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
       18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4202	100.0	833	9	US-09-844-281-1	Sequence 1, Appli
2	564.5	13.4	529	12	US-10-282-122A-45596	Sequence 45596, A
3	556.5	13.2	414	12	US-10-282-122A-45560	Sequence 45560, A
4	553.5	13.2	525	12	US-10-282-122A-46674	Sequence 46674, A
5	459	10.9	379	12	US-10-282-122A-45308	Sequence 45308, A
6	374	8.9	1222	9	US-09-137-531-15	Sequence 15, Appli
7	374	8.9	1252	9	US-09-137-531-9	Sequence 9, Appli
8	369	8.8	408	12	US-10-282-122A-46594	Sequence 46594, A
9	367.5	8.7	921	9	US-09-117-447-6	Sequence 6, Appli
10	257.5	6.1	1599	13	US-10-092-880-9	Sequence 9, Appli
11	249.5	5.9	1600	13	US-10-092-880-10	Sequence 10, Appli
12	249.5	5.9	2353	9	US-09-797-862-33	Sequence 33, Appl
13	246	5.9	1180	14	US-10-193-764-61	Sequence 61, Appl
14	246	5.9	1188	14	US-10-193-764-59	Sequence 59, Appl
15	242.5	5.8	1220	14	US-10-193-764-28	Sequence 28, Appl
16	242.5	5.8	1226	14	US-10-193-764-26	Sequence 26, Appl
17	235	5.6	1004	14	US-10-193-764-53	Sequence 53, Appl
18	235	5.6	1010	14	US-10-193-764-51	Sequence 51, Appl
19	235	5.6	2712	12	US-10-282-122A-67070	Sequence 67070, A
20	230.5	5.5	2154	12	US-10-282-122A-66695	Sequence 66695, A
21	230	5.5	1325	9	US-09-741-669-304	Sequence 304, App
22	229	5.4	969	14	US-10-193-764-32	Sequence 32, Appl
23	229	5.4	975	14	US-10-193-764-30	Sequence 30, Appl
24	227.5	5.4	1073	14	US-10-193-764-45	Sequence 45, Appl
25	227.5	5.4	1079	14	US-10-193-764-43	Sequence 43, Appl
26	223.5	5.3	852	12	US-10-282-122A-43144	Sequence 43144, A
27	222	5.3	1833	14	US-10-175-282-4	Sequence 4, Appli
28	222	5.3	1833	14	US-10-175-275-4	Sequence 4, Appli
29	222	5.3	1992	14	US-10-175-282-3	Sequence 3, Appli
30	222	5.3	1992	14	US-10-175-275-3	Sequence 3, Appli
31	222	5.3	2265	12	US-10-282-122A-45123	Sequence 45123, A
32	220.5	5.3	4327	15	US-10-369-493-10178	Sequence 10178, A
33	218.5	5.2	2122	9	US-09-813-214A-9	Sequence 9, Appli
34	217.5	5.2	1005	14	US-10-193-764-41	Sequence 41, Appli
35	217.5	5.2	1011	14	US-10-193-764-39	Sequence 39, Appl
36	217	5.2	1483	12	US-10-282-122A-51483	Sequence 51483, A
37	216.5	5.2	1095	14	US-10-193-764-65	Sequence 65, Appl
38	216.5	5.2	1536	13	US-10-092-880-2	Sequence 2, Appli
39	216.5	5.2	1536	14	US-10-193-764-63	Sequence 63, Appl
40	216	5.1	3217	12	US-10-282-122A-61210	Sequence 61210, A
41	215.5	5.1	1222	14	US-10-193-764-37	Sequence 37, Appl
42	215.5	5.1	1228	14	US-10-193-764-34	Sequence 34, Appl
43	214.5	5.1	1946	12	US-10-282-122A-62947	Sequence 62947, A
44	213	5.1	1098	9	US-09-797-862-32	Sequence 32, Appli
45	211.5	5.0	3241	9	US-09-841-786-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-09-844-281-1
; Sequence 1, Application US/09844281
; Patent No. US20020082386A1
; GENERAL INFORMATION:
; APPLICANT: Mangold, Beverly L.
; APPLICANT: Aldrich, Jennifer L.
; APPLICANT: O'Brien, Thomas
; TITLE OF INVENTION: Anthrax Specific Antibodies
; FILE REFERENCE: 38602.0003
; CURRENT APPLICATION NUMBER: US/09/844,281
; CURRENT FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/200,505
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 833
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-09-844-281-1

Query Match             100.0%;  Score 4202;  DB 9;  Length 833;
Best Local Similarity   100.0%;  Pred. No. 1.1e-290;
Matches  833;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE 60

Qy         61 NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV 120

Qy        121 NGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYGK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYGK 180
```

```
Qy   181  KDNAQAYYTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVTLGGK  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  KDNAQAYYTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVTLGGK  240

Qy   241  VAPNKDLTVKKNQSFVTKFVYKFVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLN  300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  VAPNKDLTVKKNQSFVTKFVYKFVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLN  300

Qy   301  LANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLTVSNTG  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  LANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLTVSNTG  360

Qy   361  IITVKNLDTPASAIKNVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLNKLVA  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  IITVKNLDTPASAIKNVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLNKLVA  420

Qy   421  TIAGEDKVVDPGSISIKSSNHGIIISVVNNYITAEAAGEATLIKVGDVTKDVKFKVTTDS  480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  TIAGEDKVVDPGSISIKSSNHGIIISVVNNYITAEAAGEATLIKVGDVTKDVKFKVTTDS  480

Qy   481  RKLVSVKANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVV  540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  RKLVSVKANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVV  540

Qy   541  TTDSGSIGTKTIGVTGNDVGEGTVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQY  600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  TTDSGSIGTKTIGVTGNDVGEGTVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQY  600

Qy   601  GQSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVT  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  GQSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVT  660

Qy   661  GKTPGKVDIHLTKNGATAGKATVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLELEK  720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  GKTPGKVDIHLTKNGATAGKATVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLELEK  720

Qy   721  SNLDDIVKGINLIKETQHKVRVVKSGAEQGKLYLDRNGDAVFNAGDVKLGDVTVSQTSDS  780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721  SNLDDIVKGINLIKETQHKVRVVKSGAEQGKLYLDRNGDAVFNAGDVKLGDVTVSQTSDS  780

Qy   781  ALPNFKADLYDTLTTKYTDKGTLVFKVLKDKDVITSEIGSQAVHVNVLNNPNL  833
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  ALPNFKADLYDTLTTKYTDKGTLVFKVLKDKDVITSEIGSQAVHVNVLNNPNL  833
```

RESULT 2
US-10-282-122A-45596
; Sequence 45596, Application US/10/282,122A
; Publication No. US20040029129A1
GENERAL INFORMATION:

;   APPLICANT: Wang, Liangsu
;   APPLICANT: Zamudio, Carlos
;   APPLICANT: Malone, Cheryl
;   APPLICANT: Haselbeck, Robert
;   APPLICANT: Ohlsen, Kari
;   APPLICANT: Zyskind, Judith
;   APPLICANT: Wall, Daniel
;   APPLICANT: Trawick, John
;   APPLICANT: Carr, Grant
;   APPLICANT: Yamamoto, Robert
;   APPLICANT: Forsyth, R.
;   APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20

```
Query Match         13.4%;    Score 564.5;   DB 12;   Length 529;
Best Local Similarity   32.6%;   Pred. No. 1.1e-31;
Matches  172;   Conservative   81;   Mismatches  192;   Indels  83;   Gaps  18;

Qy     1  AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE  60
          :  | : ::  | :::  | | :  | :|:   :| |::    |:|:|
Db    25  AAKKFSDVPT--WAQQSVDYLVGKKALDGKPDGTFSPSEAVDRGSAAKILAVVLGLPVDP  82

Qy    61  NAQPSFKDAKNIWSSKYIAAVEKAGVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV  120
          ::   | |::    ||  | :| |:|  :  | ::  |::||| :   | : |  |:|
Db    83  KAKPSFKDAQNHWAAPYIAAVEKAGVINGDGTGKFNPSSQINRASMASMLVQAYSLDKKI  142

Qy   121  NGELVTTFEDLLDHWGEEKANILINLGISVTGTGGKWEPNKSVSRAEAAQFIALTD----  175
          ||  | | :  |  | :  :| : | :  |  || ::|   ||   | : |:
Db   143  IGELPTQFKDLEPHWGKKQANILVALEISKGTGNGWNPEGTVTRAEAAQFIAMADQNKTS  202

Qy   176  --KKYGKKDNAQAY-------VTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIE  225
          :  |   | | :    ::: | ::|
Db   203  TSKRMYMNRNVITYHQPSLSSGITDVQ-HKPQMVEVT-----------------------E  239

Qy   226  ASTDGTSAVVT-LGGKVAPNKDLTVKVK--NQSFVTKFVYEVKKLAVEKL-TFDD----  276
          |:|   : ||   ::|
Db   240  QRADGWLKIVTSKGEKWTP---LTEKTEINEEFTI---YETASHSSKVLGTYNAQTVTV  293

Qy   277  -DRAGQAIAFKLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIFEGGEATSTTGKLAV  335
           :  ::   |:   : ::
Db   294  MEESGSWIRIRVGAGFQWVDKQNQLNPVKQB-----NFLEGK-AIIDPGHGGIDSGNVGY  347

Qy   336  GIKQGDYKVEVQV------TKRGGLTVSNTGIITVK-NLDTPASAIKNVVFALDADNDGV  388
          || || | :|:|:       | ::| ||  |  |  :
Db   348  YEKESETVLDVSLRLKKIFEQKAPFTVMFTRTDNTRPGVNSTDSLKKRVEFAQEHNGDIF  407

Qy   389  VNYGSKLSGKDFALNSQNL------VVGEKASLNKLVATIAGEDKVVDPGSISIKSSN  440
          |: |      || |::||       ::|||
Db   408  VSIHANGSAEKNGQGTETLYYQSARAKVTNPHVEDSKLLA-----QKIQDRLVAALGTKD  462

Qy   441  HGIISVVNNYITAEAAGEATLIKVGDVTKDVKVKFKVTTDSRKLVSVKA  488
          :  :|:|  | :|| :| | :  :|  |||    ||  :  |  |:  |
Db   463  RG-VKHQDLYYTRENTMPAVLTELAFVDNKSDADKIATPKQRQAAAEA  509
```

RESULT 3
US-10-282-122A-45560
; Sequence 45560, Application US/10/282,122A
; Publication No. US20040029129A1
GENERAL INFORMATION:

;   APPLICANT: Wang, Liangsu
;   APPLICANT: Zamudio, Carlos
;   APPLICANT: Malone, Cheryl
;   APPLICANT: Haselbeck, Robert
;   APPLICANT: Ohlsen, Kari

```
;     APPLICANT:  Zyskind, Judith
;     APPLICANT:  Wall, Daniel
;     APPLICANT:  Trawick, John
;     APPLICANT:  Carr, Grant
;     APPLICANT:  Yamamoto, Robert
;     APPLICANT:  Forsyth, R.
;     APPLICANT:  Xu, H.
;     TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;     FILE REFERENCE: ELITRA.034A
;     CURRENT APPLICATION NUMBER: US/10/282,122A
;     CURRENT FILING DATE:  2003-02-20
;     PRIOR APPLICATION NUMBER: 60/191,078
;     PRIOR FILING DATE:  2000-03-21
;     PRIOR APPLICATION NUMBER: 60/206,848
;     PRIOR FILING DATE:  2000-05-23
;     PRIOR APPLICATION NUMBER: 60/207,727
;     PRIOR FILING DATE:  2000-05-26
;     PRIOR APPLICATION NUMBER: 60/230,335
;     PRIOR FILING DATE:  2000-09-06
;     PRIOR APPLICATION NUMBER: 60/230,347
;     PRIOR FILING DATE:  2000-09-09
;     PRIOR APPLICATION NUMBER: 60/242,578
;     PRIOR FILING DATE:  2000-10-23
;     PRIOR APPLICATION NUMBER: 60/253,625
;     PRIOR FILING DATE:  2000-11-27
;     PRIOR APPLICATION NUMBER: 60/257,931
;     PRIOR FILING DATE:  2000-12-22
;     PRIOR APPLICATION NUMBER: 60/267,636
;     PRIOR FILING DATE:  2001-02-09
;     PRIOR APPLICATION NUMBER: 60/269,308
;     PRIOR FILING DATE:  2001-02-16
;     Remaining Prior Application data removed - See File Wrapper or PALM.
;     NUMBER OF SEQ ID NOS: 78614
;     SOFTWARE: Patentin version 3.1
;
SEQ ID NO 45560
;     LENGTH: 414
;     TYPE: PRT
;     ORGANISM: Bacillus anthracis
US-10-282-122A-45560

Query Match        13.2%;     Score 556.5;  DB 12;  Length 414;
Best Local Similarity  39.2%;   Pred. No. 2.9e-31;
Matches 145;  Conservative 54;  Mismatches 120;  Indels 51;  Gaps 14;

Qy         3 KSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENA 62
             |:||||||||||||| ||| ||| : |  |  || :| :|||:||| :||| |||| : |
Db        27 KTFPDVQTGHWAEDSINYLAEKGAVTGNEKGMFEPGKEITRAEAATMMAKILNLPIDKNA 86

Qy        63 QPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNG 122
             :||: ||: :||: |:|||||||||||| || ||:|||||:|| ||  |:|| |||:|||
Db        87 KPSYADSQKHWATPIIAAVEKAGVVKGTG-NGFEPDGKIDRVSMASLLVEAYKLESKVNG 145

Qy       123 ELVTTFEDLLD-HWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDK-KYGK 180
              | |  |:|:| |||:| |||| ||:|||||||||:|:  ||:|||| |:|  ||  |||
Db       146 TPATKFKDLETLNWGKEKANILVELGISVGTGDKWEPKKILTKAEAAQFIKKADSLKVGN 205

Qy       181 KDNAQAYVTDV--------KVSEPTKLTL-TGTGLDKLSADDVTLEGDKAVAI--EA 226
             | : |  :  |        ||| |:| |  |||| ||: |||:|||:||| ||  ::
Db       206 PLVEKVVIIDPGHGGFDPGNPGQGVEESKIVFDTSLRLQKLLEKNTPL---KALLTREEN 262

Qy       227 STDGTSAVVTLGGKV---APNKDLTVKV------KNQSFVTKFVYEVKKLAVEKLTFDD 276
             :|     : :| |:|   :|  |:|:|       |  : ::| : |:||:||||| ||
Db       263 GNPGSNKNESLANRVKFGQENNADIPVSIHANSSEKHDGHGTETYYKKSKRGETQIEK 322

Qy       277 DRAGQAIAFKLNDEKGNADVEYLNLANHDVK-----FVANNLDGSPA---------NIFE 322
             : || :|| |:|:   |  ::|: | : ::         | ||  :           :
Db       323 D--SEVLAKKIQ----NRVVEALHTRDRKIKDDHSLYVVNN-NTVPAVLTELAFIDNDID 375

Qy       323 GGEATSTTGK 332
             : :|  |||
Db       376 NGKLATESGR 385
```

```
RESULT 4
US-10-282-122A-46674
;     Sequence 46674, Application US/10282122A
;     Publication No. US20040029129A1
;     GENERAL INFORMATION:
;     APPLICANT:  Wang, Liangsu
;     APPLICANT:  Zamudio, Carlos
;     APPLICANT:  Malone, Cheryl
;     APPLICANT:  Haselbeck, Robert
;     APPLICANT:  Ohlsen, Kari
;     APPLICANT:  Zyskind, Judith
;     APPLICANT:  Wall, Daniel
;     APPLICANT:  Trawick, John
;     APPLICANT:  Carr, Grant
;     APPLICANT:  Yamamoto, Robert
;     APPLICANT:  Forsyth, R.
;     APPLICANT:  Xu, H.
;     TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;     FILE REFERENCE: ELITRA.034A
;     CURRENT APPLICATION NUMBER: US/10/282,122A
;     CURRENT FILING DATE:  2003-02-20
;     PRIOR APPLICATION NUMBER: 60/191,078
;     PRIOR FILING DATE:  2000-03-21
;     PRIOR APPLICATION NUMBER: 60/206,848
;     PRIOR FILING DATE:  2000-05-23
;     PRIOR APPLICATION NUMBER: 60/207,727
;     PRIOR FILING DATE:  2000-05-26
;     PRIOR APPLICATION NUMBER: 60/230,335
;     PRIOR FILING DATE:  2000-09-06
;     PRIOR APPLICATION NUMBER: 60/230,347
;     PRIOR FILING DATE:  2000-09-09
;     PRIOR APPLICATION NUMBER: 60/242,578
;     PRIOR FILING DATE:  2000-10-23
;     PRIOR APPLICATION NUMBER: 60/253,625
;     PRIOR FILING DATE:  2000-11-27
;     PRIOR APPLICATION NUMBER: 60/257,931
;     PRIOR FILING DATE:  2000-12-22
;     PRIOR APPLICATION NUMBER: 60/267,636
;     PRIOR FILING DATE:  2001-02-09
;     PRIOR APPLICATION NUMBER: 60/269,308
;     PRIOR FILING DATE:  2001-02-16
;     Remaining Prior Application data removed - See File Wrapper or PALM.
;     NUMBER OF SEQ ID NOS: 78614
;     SOFTWARE: PatentIn version 3.1
;
SEQ ID NO 46674
;     LENGTH: 525
;     TYPE: PRT
;     ORGANISM: Bacillus anthracis
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (296)..(296)
;     OTHER INFORMATION: X=any amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (364)..(364)
;     OTHER INFORMATION: X=any amino acid
;     FEATURE:
;     NAME/KEY: MISC_FEATURE
;     LOCATION: (493)..(493)
;     OTHER INFORMATION: X=any amino acid
US-10-282-122A-46674

Query Match        13.2%;     Score 553.5;  DB 12;  Length 525;
Best Local Similarity  31.8%;   Pred. No. 6.7e-31;
Matches 164;  Conservative 71;  Mismatches 172;  Indels 109;  Gaps 15;

Qy         1 AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE 60
              | |:||||||||||||| |||||| ::| || ||: ||::|||:||| :||| ||| |
Db        20 AAKKFSDVPT--WAQQSVDYLVGKKALDGKPDGTFSPSEAVDRGSAAKILAVVLGLPVDP 77

Qy        61 NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV 120
              |:|| |||: :||: |:|||||||| :|| || ||:| |||||:|| |||:|  ||||
Db        78 KAKPSFKDAQNHWAAPYIAAVEKAGVINGDGTGKFNPSSQINRASMASMLVQAYSLDKKI 137
```

```
;   SOFTWARE: PatentIn version 3.1
;   SEQ ID NO 45308
;   LENGTH: 379
;   TYPE: PRT
;   ORGANISM: Bacillus anthracis
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (227)..(227)
;   OTHER INFORMATION: X=any amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (229)..(229)
;   OTHER INFORMATION: X=any amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (255)..(255)
;   OTHER INFORMATION: X=any amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (304)..(304)
;   OTHER INFORMATION: X=any amino acid
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (369)..(369)
;   OTHER INFORMATION: X=any amino acid
US-10-282-122A-45308

Query Match             10.9%;   Score 459;  DB 12;   Length 379;
Best Local Similarity   32.5%;   Pred. No. 2.2e-24;
Matches  124;  Conservative  49;  Mismatches  110;  Indels  98;  Gaps  13;

Qy      5  FPDVPAGHWAEGSINYLVDVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQP  64
              ||||| :| ||||     :: ||| |||::         ::::|| : | :  :
Db     24  FPDVPA--WADKSVTYLVDKQVLSGYPDGTFGSSDTLDRASAATIMTTALGIHIDLNAKP  81

Qy     65  SFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGEL  124
              |:||::|  || :|||    :: | :||||  ||| |||::| ::|| | :||::
Db     82  SFKDSQNHWGTPYMPAAEKAGSIKVEGKGIFNPSGKVTRAAMATMLVNAYKLQNKNTSNG  141

Qy    125  VTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEBAAQFIALTDKKYGKKDNA  184
              |||||||||||||||||||||:|||||||:|||||||||||::|||:||||||||||
Db    142  QSKFEFKGHWGGKIPNTLIGFEISVGTDNGWQPNKFITRAEEAAQLTAKTD--------  192

Qy    185  QAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVTLGGKVAPN  244
              :|: :  : ||  |::
Db    193  --MLQYSHSNPLE-----------------------------NKTIIIDPGHGGED----PG  219

Qy    245  KDLTVKVKNQSFVTKFVYEVKKLAVEKLT--------FDDDRAGQAIAFKLNDEKGNADV  296
              || ||:  || ::|| ||  :||  :
Db    220  KD-TKGLPXKXIVLDTSLRLQKL-LEKHTPFTVLLNSXSDTRTGHGSKSSL-QERG----  272

Qy    297  EYLNLANHDVKFVANNLDGSPANIPEG-GEATST-----------------TGKLA  334
              ::    |: : ||    ::  |
Db    273  KPAKTKQGDILIMGH------ANAFNGNGKR7ETYYYXSSKSEKTNPHVEKKPGFPGKIQ  326

Qy    335  V----------GIKQGDYKV  344
              |         ||:|
Db    327  TRLVDALQTRDRGVKHGDLHV  347
```

```
Qy    121  NGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEBAAQFIALTDKKYGK  180
              |:||| |||||||||||||||:||||||| ::|||||||||||::|||:|||||||||
Db    138  IGELPTQFKDLPPHWGKKQANILVALEISKGTGNGWNPEGTVTRAEAAQFIAMAEKKIKQ--  197

Qy    181  KDNAQAY----------VTDV------KVSEPTKLTLTGTGLDKLSA--------  211
              :  :  :        :  :    |||
Db    198  STSKRMYMTRNVITYHHPSLSSGITDVHHKPQNGKVTEQ-----RADGWVKQLTSKREKW  252

Qy    212  DDVTLEGDKAVAIEASTDGTSAVVTLGGKVAPNKDLTVKVKNQSFVTKFVYEVKKLAVEK  271
              :         |||  |
Db    253  -----ASHSSKVLGTYNAQTVTVMEKEWXLIRI--  300

Qy    272  LTFDDDRAGQAIAPKLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIPEGGEATSTTG  331
              |  |        |||:||  |
Db    301  ------RVGAGFQW--------VDKNQLNPVKQE-----NFLEGK-AIIIDPGHGGIDSG  340

Qy    332  KLAVGIKQGDYKVEVQ------VTKRGGLTVSNTGIITVK-NLDTPASAIKNVVFALDAD  384
              |  |::     :: | :   | |:| :    :
Db    341  NVGYYEKESDTVLDVSLRLKKIXAKAPFTVMFTRTDNTRPGVNSTDSLKKRVEFAQEHN  400

Qy    385  NDGVVNYGSKLSGKDFALNSQNL----------------VVGEKASLNKLVATIAGED  426
              :   :   ::|| |      |         ::|  |  :|
Db    401  GDIFVSIHANGSAEKNGQGTETLYYQSARAKVTNPHVEDSKLLAQKIS-DRLVAALGTKD  459

Qy    427  KVVDPGSISIKSSNHGIISVVNNYITAEAAGEATLT  462
              |   |       ::||
Db    460  RGVKHQDL------------------YVTRENTLPAVLT  480
```

```
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/137,531
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/682,517
;     FILING DATE:
;   INFORMATION FOR SEQ ID NO: 15:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 1222 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
;       MOLECULE TYPE: protein
US-09-137-531-15
```

```
Query Match       8.9%;  Score 374;  DB 9;  Length 1222;
Best Local Similarity  23.3%;  Pred. No. 1.5e-17;
Matches 224;  Conservative 130;  Mismatches 345;  Indels 264;  Gaps  47;

Qy   13 WAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQPSFKDAK-N  71
Db   12 YAKEAVQALVDQGVIQGDTNGNFNPLNTVTRAQAAEIFTKALEL--EANGDVNFKDVKAG  69

Qy   72 IWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGELVTTFEDL 131
Db   70 AWYYNSIAAVVANGIFEGVSATEFAPNKSLTRSEAAKILVEAFGLEGEAD---LSEFADA 126

Qy  132 --LDHWGEEKANILINLGISVGT-GGKWEPNKSVSRAE-AAQFIALTDKKYGKKDNAQAY 187
Db  127 SQVKPWAKKYLEIAVANGIFEGTDANKLNPNNSITRQDFALVFKRTVDKVEGETPEEAAF 186

Qy  188 V-----TDVKVS--EPTKLTLTGLDKLSADDVTL-EGDKAVAI---EASTDGTSAVV    235
Db  187 VKAINNTTVEVTFEEEVTNVQALNFKIEGLEIKNASVKQTNKKVVVLTTEAQTADKEYVL 246

Qy  236 TLGGK------------VAPNKDLTVKVKNQSFVTKFVYEVKKLLAVEKLTFDDDRAGQAIAF 285
Db  247 TLDGETIGGFKGVAAVVPTK---VELVSSAVQGKLGQEVKVQAKVTVAEGQSKAGIPVTF 303

Qy  286 -----------------KLNDEKGNADVEYLNLANHDVKFVA----NNLDGSPANIFEGG 324
Db  304 TVPGNNNDGVVPTLTGEALTNEEGIATYSYTRYKEGTDEVTAYATGDRSKFSLGYVFWGV 363

Qy  325 EATSTTGKLAVG---IKQG---DYKVEVQVTKRGGLTVS--NTGIITVKNL--DTPASAI 374
Db  364 DTILSVEEVTTGASVNNGANKTYKVTYVTYKNPKTGKPEANKTFNVGFVENMNVTSDKVANAT 423

Qy  375 KNVVPALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLNKLVAATIAGEDKVVDDGSI 434
Db  424 VNGVKALQLSN------GTALDAAQITTDSK------------------GEATFVSGTNAAVTPVVY 467

Qy  435 SIKSSNHGIISVVNNYITAE----------------AAGEATLTIKV------GDV-----TKDV 472
Db  468 DLHSTNN---STSNKKYSASALQCTASKVTFAALQAEYTIELTRADNAGEVAAIGATNGR 524

Qy  473 KFKVTTDSR------KLVSVKANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEV 525
Db  525 EYKVIVKDKAGNLAKNEIVNVAFNEDKDRVISTVT-NAKFVDTD------------ 567

Qy  526 LPKTGVVAEGG----LDVVTTDSGS----IGTKT---------IGVTGNDVGEGTV-- 564
Db  568 -PDTAVYFTGDKAKQISVKTNDKGBATFVIGSDTVNDYATPIAWIDINTSDAKQGDLDEG 626

Qy  565 -------HFQ---NGNG------ATLGSLYVNVTEGNVAPKNFELVSKVGQ--- 599
Db  627 EPKAVAPISYFQAPYLDGSAIKAYKKSDLNKAVTKFDGSETAVFAAEIVNQSGKKVTGTS 686

Qy  600 -YGQ-SPDTKLDLNV----STTVEYQLSKYTSDRVYSDPENLEGYEVESKNL 645
Db  687 IKFATYTYNTGANDIKVDNQVISPNRSYTVTYEATLSSTGTVITPAKNLEVTSVDGKTT 746

Qy  646 AVADAKIVGNKVVVTG---KTPGKVDIHLTKNGATAG-KATVEIVQETIAIKSVNFKPVQ 701
```

```
Db  747 AV-------KVIATGIAVNTDGK-DYAFTAKEATATFATNEVPN------SYTGVA 789
Qy  702 TE-NFVEKKINIGTVLELEKSNLDDIVKGINLTKETQHKVRVVKSGAEQGKLJ--YLDRNG 758
Db  790 TQFNTADSGSNSNSIWFAGKN---------------------PVKYAGVSGKTYKYFGANG 829
Qy  759 DAVFNAGDVKLGDVTVSQTSDSALNFKADLYDTLTTKYTDKG---TLVFKVLKDKDVIT 815
Db  830 NEVFGEA----------------------AWEALLTQYATEGQKVTISYNV--DGDTVT 864

Qy  816 SEI 818
Db  865 FKV 867
```

```
RESULT 7
US-09-137-531-9
; Sequence 9, Application US/09137531
; Patent No. US2002004816A1
; GENERAL INFORMATION:
;  APPLICANT:
;  TITLE OF INVENTION:  Expression of surface layer proteins
;  NUMBER OF SEQUENCES: 25
;  COMPUTER READABLE FORM:
;    MEDIUM TYPE: Floppy disk
;    COMPUTER:  IBM PC compatible
;    OPERATING SYSTEM:  PC-DOS/MS-DOS
;    SOFTWARE:  Patentin Release #1.0, Version #1.30 (EPO)
;  CURRENT APPLICATION DATA:
;    APPLICATION NUMBER: US/09/137,531
;    FILING DATE:
;    CLASSIFICATION:
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER:  08/682,517
;    FILING DATE:
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1252 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;      MOLECULE TYPE:  protein
US-09-137-531-9
```

```
Query Match       8.9%;  Score 374;  DB 9;  Length 1252;
Best Local Similarity  23.3%;  Pred. No. 1.6e-17;
Matches 224;  Conservative 130;  Mismatches 345;  Indels 264;  Gaps  47;

Qy   13 WAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQPSFKDAK-N  71
Db   12 YAKEAVQALVDQGVIQGDTNGNFNPLNTVTRAQAAEIFTKALEL--EANGDVNFKDVKAG  99

Qy   72 IWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGELVTTFEDL 131
Db  100 AWYYNSIAAVVANGIFEGVSATEFAPNKSLTRSEAAKILVEAFGLEGEAD---LSEFADA 156

Qy  132 --LDHWGEEKANILINLGISVGT-GGKWEPNKSVSRAE-AAQFIALTDKKYGKKDNAQAY 187
Db  157 SQVKPWAKKYLEIAVANGIFEGTDANKLNPNNSITRQDFALVFKRTVDKVEGETPEEAAF 216

Qy  188 V-----TDVKVS--EPTKLTLTGLDKLSADDVTL-EGDKAVAI---EASTDGTSAVV    235
Db  217 VKAINNTTVEVTFEEEVTNVQALNFKIEGLEIKNASVKQTNKKVVVLTTEAQTADKEYVL 276

Qy  236 TLGGK------------VAPNKDLTVKVKNQSFVTKFVYEVKKLLAVEKLTFDDDRAGQAIAF 285
Db  277 TLDGETIGGFKGVAAVVPTK---VELVSSAVQGKLGQEVKVQAKVTVAEGQSKAGIPVTF 333

Qy  286 -----------------KLNDEKGNADVEYLNLANHDVKFVA----NNLDGSPANIFEGG 324
Db  334 TVPGNNNDGVVPTLTGEALTNEEGIATYSYTRYKEGTDEVTAYATGDRSKFSLGYVFWGV 393

Qy  325 EATSTTGKLAVG---IKQG---DYKVEVQVTKRGGLTVS--NTGIITVKNL--DTPASAI 374
```

```
Db   394  DTILSVEEVTTGASVNNGANKTYKVTYKNPKTGKPEANKTFNVGFVENMNVTSDKVANAT  453
                                                                      |
Qy   375  KNVVFALDADNDGVNYGSKLSGKDFALNSQNLVVGEKASLNKLVATIAGEDKVVDPGSI  434
               :  ||  :|    ||  |       :||:  :
Db   454  VNGVKALQLSN------GTALDAAQITTDSK------------------  497
                 || |                      |  |
Qy   435  SIKSSNHGIISVVNNYITAE---------AAGEAILTIKV-----GDV------TKDV  472
           : |:|                      ||   :|        :
Db   498  DLHSTNN---STSNKKYSASALQTTASKVTPAALQAEYTIELTRADNAGEVAAIGATNGR  554
              |  |
Qy   473  KFKVTTDSR-------KLVSVKANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEV  525
           :|:        :  |:|   | ||||||:   ||
Db   555  EYKVIVKDKAGNLAKNEIVNVAFNEDKDRVISTVT-NAKFVDTD---------------  597
              |  |  |                  |
Qy   526  LPKIGVAEGG----LDVVTTDSGS----IGTKT---------IGVTGNDVGEGTV---  564
            ||  |         ||:    |:  :||  |        ::
Db   598  -PDTAVYFTGDKAKQISVKTNDKGEATFVIGSDTVNDYATPIAWIDINTSDAKQGDLDEG  656
              | |  |   |      |  | |  :    |       ||
Qy   565  --------HFQ-----NGNG------ATLGSLYVNVTEGNVAFKNFELVSKVGQ-----  599
                                           :||         |:|:
Db   657  EPKAVAP:ISYFQAPYLDGSAIKAYKKSDLNKAVTKFPDGSETAVFAAELVNQSGKKVTGTS  716
              |  :         :| ||          ||    |   :||  |||  ||
Qy   600  ----YGQ-SPDTKLDLNV----STTVEYQLSKYTSDRVYSDPENLEGYEVESKNL  645
                |                 :   |     :  :  :  ||  |
Db   717  IKKATYTINTGANDIKVDNQVISPNRSYTVTYEATLSSTGTVITPAKNLEVTSVDGKTT  776
              |                  |  |  || |    |   | |
Qy   646  AVADAKIVGNKVVVTG---KTPGKVDIHLTKNGATAG-KATVEIVQETIAIKSVNFKPVQ  701
                 | |          ||  :     |:       |        |
Db   777  AV------KVIATGIAVNTDGK-DYAFTAKEATATFTATNEVPN-------SYTGVA  819
                      |   |             :                   |
Qy   702  TE-NFVEKKINIGTVLELEKSNLDDIVKGINLTKETQHKVRVVKSGAEQGKL--YLDRNG  758
                       |:   :           |                   :||
Db   820  TQFNTADSGSNSNSIWFAGKN--------------PVKYAGVSGKTYKYFGANG  859
                |:                              :   |      :|
Qy   759  DAVFNAGDVKLGDVTVSQTSDSALPNFKADLYDTLTTKYTDKG---TLVFKVLKDKDVIT  815
              |  ||                 :| :|    |: :|  :|||   | :
Db   860  NEVFGEA---------AWEALLTQYATESQKVTISYNV--DGDTVT  894
              :                   :    ||
Qy   816  SEI  818
           ::
Db   895  FKV  897
```

```
RESULT 8
US-10-282-122A-46594
; Sequence 46594, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT: Zamudio, Carlos
;  APPLICANT: Malone, Cheryl
;  APPLICANT: Haselbeck, Robert
;  APPLICANT: Ohlsen, Kari
;  APPLICANT: Zyskind, Judith
;  APPLICANT: Wall, Daniel
;  APPLICANT: Trawick, John
;  APPLICANT: Carr, Grant
;  APPLICANT: Yamamoto, Robert
;  APPLICANT: Forsyth, R.
;  APPLICANT: Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE: 2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
```

```
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 46594
;  LENGTH: 408
;  TYPE: PRT
;  ORGANISM: Bacillus anthracis
;  FEATURE:
;  NAME/KEY: MISC_FEATURE
;  LOCATION: (161)..(161)
;  OTHER INFORMATION: X=any amino acid
;  FEATURE:
;  NAME/KEY: MISC_FEATURE
;  LOCATION: (165)..(165)
;  OTHER INFORMATION: X=any amino acid
;  FEATURE:
;  NAME/KEY: MISC_FEATURE
;  LOCATION: (180)..(180)
;  OTHER INFORMATION: X=any amino acid
;  FEATURE:
;  NAME/KEY: MISC_FEATURE
;  LOCATION: (184)..(184)
;  OTHER INFORMATION: X=any amino acid
;  FEATURE:
;  NAME/KEY: MISC_FEATURE
;  LOCATION: (187)..(185)
;  OTHER INFORMATION: X=any amino acid
;  FEATURE:
;  NAME/KEY: MISC_FEATURE
;  LOCATION: (303)..(303)
;  OTHER INFORMATION: X=any amino acid
US-10-282-122A-46594

Query Match        8.8%;  Score 369;  DB 12;  Length 408;
Best Local Similarity   31.0%;  Pred. No. 6.5e-18;
Matches  148;  Conservative  57;  Mismatches  172;  Indels  100;  Gaps  21;
```

```
Qy     3  KSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENA   62
          | ||||| | || :|  :|::||||||||||| : :|||||:|||||||||:|||:||
Db    22  KFFPDVQQGPGAEYSINYLAEKGAVTGNEKGMFEPGKEITRAEAATMMAKILNLPIDKNA   81
Qy    63  QPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNG  122
          ||: ||  :|:| | :  |::||:|||||||| :||:|||| ::|||:|:||||||||
Db    82  KPSYADSQKHWATPIIAAVEKAGVVKGTG-NGFEPDGKIDRVSMASLLVEAYKLESKVNR  140
Qy   123  -ELVTTFEDLLDH-WGEEKANIL-INLGISVGTGGKWEPNKS-VSRAEAAQFIALTDKKY  178
           :|||             :||    ||| :||:|||||: : |
Db   141  FPQPINFKSFKNHKRGKEKRXYFKKSWVFFVGPGGQORGPXENFXXKAEPVHFIKRPD---  197
Qy   179  GKKDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVTLG  238
                                                        |
Db   198  ---------SLKVGNPLV--------------------------------EKVVIIDPGH---------G  217
Qy   239  GKVAPNKDLTVKVKNQSFVTKFVYKKLAVEKLTFDDDRAGQAIAFKLNDEKGN-ADVE  297
              :|| ||                  ::
Db   218  GFDPGNPGQGYB------ESKIVFDT-SLRIQKLL--EKNTPLKALLTREENGNPGSNK  267
Qy   298  YLNLANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLTVS  357
              || |   :                  ::    |:   ||
Db   268  NESLANR-VKFGQEN---NADIFVSIHANSSEKHDGHGFXNYYYK----KSKRGEETQI  318
Qy   358  NTGIITVKNLDTPASAIKN-VVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLN  416
```

```
Qy    446  VVNN----YITAEA--AGEATLTIKVGDVTK--------DVK------FKVTTDSRKLVSVK  487
                     :: ::  .:   :: .::::.:::          :::       :: :::||::::
Db    461  AAINGSELLVTANAGQSGKASFEVTLKDNTKRFTVDVKKDPVLQDIKVDATSVKLSDEA  520

Qy    488  ANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAIKEVLPKTGVVAEGGLDVVTTDSGSI  547
                |||:::.||:      :::    ::::::.                 :::  ||:::: |||::
Db    521  VGGGEVEGVNQKTIKVSAV--DQYG---------KEI--KFGT--KGKVTVTNTEGL  563

Qy    548  GTKTIGVTGNDVGEGTVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQYGQSPDTK  607
                :: :: ||||    ::::::|:   ::|:|:.:   :.:::.           :|:
Db    564  VIKNVN------SDNTIDFDSGNSAT-------DQFVVVA-------TK  592

Qy    608  LDLNVSTTVEYQLSKYTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKV  667
                |:|  ::|::| :::::.  ::::.|::::.| ::           |:|
Db    593  -DKIVNGKVEVKYFQNASD--TTPTSTKTITVNVVNVK-ADATPVGLLDIV---APSKI  643

Qy    668  DIHLTKNGATAGKATVEIVQ-ETIAIKSVN----FKPVQTE--NFVEKKINIGTV  715
                |: |:|:.| :::::|:||:::             :::.:: ::: ::
Db    644  DVNAPN---TASTADVDFINFESVEIYTLDSNGRRQKKVTPTATTLVGTKKKKKVN-GNV  699

Qy    716  LELEKSNLDDIVKGINLTKETQHKVRVVKSGAE------QGKLYLDRNGDAVFNAGDVKL  769
                ||::|:       ||:: :.:::::.:|:|:.                .:|:::::::|
Db    700  LQFKGNE------ELTLSTSSSTGNVDGTAEGMTKRIPGK-YI--NSASVPASATVAT  748

Qy    770  GDVTVS-QTSDSALPNFKADLYDFLTTKYTDKGTLVFFKVLKDKDV  813
                 ::::: :.:. ::|:      :|| :::.||
Db    749  SPVTVKLNSSDNDL-----TPEELIFGVIDPTQLV---KDEDI  783
```

```
Query Match       6.1%;   Score 257.5;  DB 13;  Length 1599;
Best Local Similarity  22.0%;   Pred. No. 4.6e-09;
Matches 183;  Conservative 104;  Mismatches 328;  Indels 215;  Gaps 38;

Qy     16  GSI-----NYLVDKGA----------ITGKPDGTYGPTESIDRASAAVIFTKILN--L   56
                |||       || |||            ::||.::|:|  ::.:::|| :::::::
Db    891  GSIINIAGNLTVSKGANLQAITNYTFNVAGSFDNGASNISTARGGAK--FKDINNTSSL  948

Qy     57  PVDENAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNL  116
                                :||.::|:|:|:|:::||.: .::.::.:::|:::::::::|:
Db    949  NITTNSDTTYR----------TIIKGNISNKSGDLNI----IDKKSDAEIQIGG-NI  990

Qy    117  KDKVNGELVTTFEDLLDHWGEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDK  176
                |:|:::||.:|:|::||:|:|||:|||||||:|:|||||||:|||:|:|||||||||:|
Db    991  SQK-EGNLTIS----------SDKVNITNQITIKAGVEG----------1018

Qy    177  KYGKKDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVT  236
                ||||::::|||||||||||||||||:.||:|| ::|:|:|::||:::::|:.
Db   1019  --GRSDSSEA------ENANLTIQTKEL-KL-AGDLNISGFNKAEITAKNGSDLTIGN  1066
```

```
Query Match       8.7%;   Score 367.5;  DB 9;  Length 921;
Best Local Similarity  24.9%;   Pred. No. 2.9e-17;
Matches 220;  Conservative 121;  Mismatches 335;  Indels 209;  Gaps 47;

Qy      4  SFPDVPAGHAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQ   63
                ||||:|.|.:|:|||||:||::|||:|::||.|:|||||||:::::::|||:||||||||
Db     33  SFTDVAPQY--KDAIDFLVSTGATKGKTETKFGVIDEITRLDAAVILARVLKLDVDNAKD   90

Qy     64  PSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASPASMLVSAYNLKDKVNGE  123
                |||:||:|||:|:|: ::||:.:|||.:::::||||:||||.|:||:|::::|||.:|:
Db     91  AGFTDVPKD-RAKYVNALVEAGVLNGKAPGKFGAYDPLTRVEMAKIIANRYKLK--ADD  146

Qy    124  LVTTFEDLLDHWGE---------EKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTD  175
                ||:||::||.|:|::         :|:|||||:||||||||||||:||:|:|||||||||:
Db    147  VKLPFFDVNDTWAPVVKALYKEVTKVKRLKHQQASVHT------KNITLRDFAQFVY---  196

Qy    176  KKYGKKDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVT-LEGDKAVAIEAST---DGT  231
                |::||:|:|:|||:|:|||:|:::::|::||.:| .:|:|| ::|:|||||||:
Db    197  ---RAVNINAVPEIEVTAVNSTTVKVTFNTQIADVDFTNFAIDNGLTVTKATLSRDKK  252

Qy    232  SAVVTLGGKVAPNKDLTVK---VKN------QSFVTKFVVEVKKLAVEKLTFDDDRAGQA  282
                |:|||.|.| ::|:|||| :|:      :||||:|.|:::::||.::|:|:||||:.||:
Db    253  SVEVVNKPFTRNQEYTITATGIKNLKGETAKELTGKFVNSVQDAVTVALNNSSLKVGEE  312

Qy    283  IAFKLNDEKGNADVEYLNLANHDVKFVANNLDGSPANI--FEGGEATSTTGKLAVGIKQG  340
                ||:||::|:|:||:| |::. ::||.:::|:|: :|:::  .::.:::::.|:::||:|:|
Db    313  SGLTVKDQDGK--------DVVGAKVELTSSNINIVVVSSGEVSVSAAKV-TAVKPG  360

Qy    341  DYKVEVQVTKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDF  400
                |||::||.:|:: |:|:||:| ::||:||.|::::.:|:|:||||||.:|:|.|:|:.|:
Db    361  TADVTAKVTLPDGVVLTNTFKVTVTEVPV---QVQNQGFTL-VDN------LSNA---  405

Qy    401  ALNSQNLVVGEKASLNKLVATIAGEDKVV------------DPGS----ISIKSSNHGIIS  445
                ||:|.|:.::|::|::||.|:::.||::.             :::: ::|||:.:|::|||:
Db    406  ---PQNTVAFNKA--EKVTSMFAGETKTVAMYDTKNGDPETKPVDFKDATVRSLNPIIAT  460
```

```
Db    319  E------KDSEVLAKKIQNRVVEALHTRDRKI--------KD----DHSLYVVNNNTVP  359
                 :: ||  ::::.::.|::|:.|:::::|:|:        |:    :::|:::::||
Qy    417  KLVATIAGEDKVVDPGSISIKSSNHGIISVNNYTTABAAGEATLTI--KVGDVTK  470
                ::::   ::   |:|||:::: ||:.  :|.::           :: ||:::||
Db    360  AVLTELAFIDNDIDNGKLATESGRQ--------IAEBAVYAGILDYYEWKGFDVSK  407
```

```
Qy   237 LGGKVAPNKDLTV-KVKNQSFVT----KFVYEVKKLAVEKLTFDDDRAGQAIAFKLNDE 290
Db  1067 ASGGNADAKKVTFDKVKDSKISTDGHNVTLNSEVKTSNGSSNAGNDNSTGLTISAK---- 1122
Qy   291 KGNADVEYLNLANHDVKFVANNLDGSPANIF--EGGEATSTTGKLAVGIKQGDYKVEVQV 348
Db  1123 ---DV---TVNNNVTSHKTINISAAAGNVTTKEGTTINATTGSVEVTAQNG------- 1167
Qy   349 TKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSCNLV 408
Db  1168 TIKGNITTSQNVTVTATENLVTTENAVINA------TSGTVNISTKTGDIKGGIESTSGN 1220
Qy   409 VGEKASLNKL-VATIAGEDKVVDPGSISIKSSNHGIISVVNNYIITAEAAGEATLTIKVGD 467
Db  1221 VNITASGNTLKVSNITGQDVTV------TADAGALTTTAGSTISATTGNANITTKTGD 1272
Qy   468 VTKDVK---------FKVTTDSRKLVSVKANPDKLQVVQNKTLPVT--FVTTDQY 511
Db  1273 INGKVESSSGSVTLVATGATLAVGNISGNTVTITADSGKLTSTVGSTINGTNSVTTSSQS 1332
Qy   512 GDPGANTAAIKEVLPKTG-------VVAEGGLDVVTTDSGSIGTKIGVTGNDVGEG 562
Db  1333 GDIEGTISGNTVNVTASTGDLTIGNSAKVEAKNGAATLTAESGKLTTQ----TGSSI--- 1385
Qy   563 TVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQYGQSPDTKLDLNVST-TVEYQLS 621
Db  1386 ---TSSNQTTLTAKDSSIAGNINAANVTL-NTTGTLTTTGDSKINATSGTLTINAKDA 1440
Qy   622 KY----TSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVDIH---LTKN 674
Db  1441 KLDGAASGDRTVVNATNASG--------SGN--VTAKTSSSVNITGDLNTIN 1482
Qy   675 G----ATAGKATVEIVQETAIKSVNFKPVQTENFVEKKINIGTVLELEKSNLDDIVK-G 729
Db  1483 GLNIISENGRNTVRLRGKEIDVKYIQPGVASVEEVIEAKRVLEKVKDLSDEERETLAKLG 1542
Qy   730 IN-------LTKETQHKVRVVKSG---AEQGKLYLDRNGDAVFNAGD 766
Db  1543 VSAVRFVEPNNAITVNTQNEFTTKPSSQVTISEGK-------ACFSSGN 1584
```

RESULT 11
US-10-092-880-10
; Sequence 10, Application US/10092880
; Publication No. US20020164354A1
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
; TITLE OF INVENTION: HAEMOPHILUS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/092,880
; CURRENT APPLICATION FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: 09/155,614
; PRIOR APPLICATION FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: PCT/US97/04707
; PRIOR APPLICATION FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 1600
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-092-880-10

Query Match   5.9%;   Score 249.5;   DB 13;   Length 1600;
Best Local Similarity   22.2%;   Pred. No. 1.7e-08;
Matches   185;   Conservative 121;   Mismatches 346;   Indels 183;   Gaps 37;

```
Qy    16 GSINYLVDKGAITGKPDGTYGPTE---------SIDRASAAVIFTKILNLPV 58
```

```
Db   850 GNIN-ITNKANVTLQADTSNSNTGLKKRTLTLGNISVEGNLSLTGANANIVGN--LSIAE 906
Qy    59 DENAQPSFKDAKNI--WSSKYIAAVEKAGVKGDGKENFYPEGKIDRASFASMLVSAYN 115
Db   907 DSTFKGBASDNLNITGTFTNNGTANINIKGVVK-------LGDINNKGGLNITTNASG 957
Qy   116 L-KDKVNGELVTTFEDLLDHWGEEKANILNLGISVGTGGKWEPNKSVSRAEAAQFIALT 174
Db   958 TQKTIINGNITNEKGDL--NIKNIKADABIQIG---GNISQXEGNLTISSDKVNITNQIT 1012
Qy   175 DK--KYGKKDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGT 231
Db  1013 IKAGVEGGRSDSSEA----ENANLTQTKEL-KL-AGDLNISGFNKAEITAKNGSD 1062
Qy   232 SAVVTLGGKVAPNKDLTV-KVKNQSFVT----KFVYEVKKLAVEKLTFDDDRAGQAIAF 285
Db  1063 LTIGNASGGNADAKKVTFDKVKDSKISTDGHNVTLNSEVKTSNGSSNAGNDNSTGLTISA 1122
Qy   286 KLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIF--EGGEATSTTGKLAVGIKQGDYK 343
Db  1123 K------DV--TVNNNVTSHKTINISAAAGNVTTKEGTTINATTGSVEVTAQNG--- 1168
Qy   344 VEVQVTKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALN 403
Db  1169 ----TIKGNITTSQNVTVTATENLVTTENAVINA------TSGTVNISTKTGDIKGGIE 1216
Qy   404 SONLVVGEKASLNKL-VATIAGEDKVVDPGSISIKSSNHGIISVVNNYITAEAAGEATLT 462
Db  1217 STSGNVNITASGNTLKVSNITGQDVTV------TADAGALTTTAGSTISATTGNANIT 1268
Qy   463 IKVGDVTKDVK--------FKVTTDSRKLVSVKANPDKLQVVQNKTLPVT--FV 506
Db  1269 TKTGDINGKVESSSGSVTLVATGATLAVGNISGNTVTITADSGKLTSTVGSTINGTNSVT 1328
Qy   507 TTDQYGDPFGANTAAIKEVLPKTG------VVAEGGLDVVTTDSGSIGTKTIGVTGN 557
Db  1329 TSSQSGDIEGTISGNTVNVTASTGDLTIGNSAKVEAKNGAATLTAESGKLTTQ----TGS 1384
Qy   558 DVGEGTVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQYGQSPDTKLDLNVST-TV 616
Db  1385 SI-------TSSNGQTTLTAKDSSIAGNINAANVTL-NTTGTLTTTGDSKINATSGTLTI 1436
Qy   617 EYQLSKY---TSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVDIH-- 670
Db  1437 NAKDAKLDGAASGDRTVVNATNASG--------SGN---VTAKTSSSVNITGD 1478
Qy   671 -LTKNG----ATAGKATVEIVQETAIKSVNFKPVQTENFVEKKINIGTVLELEKSNLDD 725
Db  1479 LNTINGLNIISENGRNTVRLRGKEIDVKYIQPGVASVEEVIEAKRVLEKVKDLSDEERET 1538
Qy   726 IVK-GIN-------LTKETQHKVRVVKSG---AEQGKLYLDRNGDAVFNAGD 766
Db  1539 LAKLGVSAVRFVEPNNAITVNTQNEFTTKPSSQVTISEGK-------ACFSSGN 1585
```

RESULT 12
US-09-797-862-33
; Sequence 33, Application US/09797862
; Patent No. US2002010227GA1
; GENERAL INFORMATION:
; APPLICANT: PEAK, IAN RICHARD ANSELM
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: MOXON, E. RICHARD
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0134
; CURRENT APPLICATION NUMBER: US/09/797,862
; CURRENT APPLICATION FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR APPLICATION FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR APPLICATION FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1

```
; SEQ ID NO 33
;  LENGTH: 2353
;  TYPE: PRT
;  ORGANISM: Haemophilus influenzae
US-09-797-862-33

Query Match             5.9%;  Score 249.5; DB 9;  Length 2353;
Best Local Similarity  22.0%;  Pred. No. 3.1e-08;
Matches  201; Conservative 107; Mismatches 335; Indels 269; Gaps  45;

Qy     23  DKGAITGKPDGTYGPTESIDRASAAV--IFTKILNLPVDENAQPSFKDAKNIWSSKYIAA   80
               :||::    |||     : ||::    |  :     :|       :||   ::|
Db    705  DRGKVTVK-DAT---ANDADKKVATVKDVATAI------NSAATFVKTENLITS-----   748

Qy     81  VEKAGVVKGDGKENFYPEGKIDRASFASMLV--SAYNLKDKVNGELVTTFEDLLDHWGEE  138
                                     |:     :|  :    ||::|: :  : :
Db    749  -------IDEDNPTDNGKDDALKAGDTLITFKAGKNLKVKRDGKNITF--DLAKNLEVK   797

Qy    139  KANI--LINLGISVGTGGKWEPNK--SVSRAEAAQFIALTDKKYGKKDNAQAYVTDVKVS  194
               ||     : || |   :||        |:|:      :|::||:||:   :::|::
Db    798  TAKVSDTLTIGGNTPTGGTTATPKVNITSTADGLNFAKETADASGSKN---VYLKGI--  851

Qy    195  EPTKLTLTGTGLDKLSADDVTLEGDK---AVAIEASTDGTSAVVTLGGKVAPNKDLTVKV  251
           |   ::|:| | :: ||| :|:::       |:|   :   :      : :     :::
Db    852  -ATTLTEPSAGA-KSSHVDLNVDATKKSNAASIE----DVLRAGWNIQGNGN------  897

Qy    252  KNQSFVTKPFVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLNLANHDVKFVAN  311
               ::      | |:  :|         |:  ||   |  ::          :   :|:
Db    898  -NVDYVATY---------------DTVNFTDDSTGTTTV---------------TVTQ  924

Qy    312  NLDGSPANIFEGGEAT--STTGKLAVGIKQGDYKVEVQVTKRGGLTVS-------NTGI  361
           ||           ::    :|:| |:       :|:::|    |   |        |:
Db    925  KADGKGADVKIGAKTSVIKDHNGKLFTGKDLKD------ANNGATVSEDDGKDTGTGL  976

Qy    362  ITVKNL----------------DTPASAIKNVVFALDADNDGVVNYGSKLSGKD-----  399
           : | :                  :|: |  :: |:         :  :: |:
Db    977  VTAKTVIDAVNKSGWRVTGEGATAETGATAV------NAGNAETVTSGTSVNFKNGNAT 1029

Qy    400  -FALNSQNLVVGEKASLN----------------KLVA-----TIAGEDKVVDPGSISIKSSN--  440
            | |:  ::|| :|:               |:|      : || :   :::: ||
Db   1030  TATVSKDNGNINVKYDVNVGDGLKIIGDDKKIVADTTTLIVTGGKVSVPAGANSVNNNKKL 1089

Qy    441  ---HGIISVNN----------YITAEAAGEATLTIKVGDVTKDVKFKVTTDSRKLVSVK-  487
              | ::||:           |:||||| ||:: |::||: |:|||||     | :|:|
Db   1090  VNAEGLATALNNLSWTAKADKYADGESEGETDQEVKAGD-------KVTFKAGKNLKVKQ 1142

Qy    488  ANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTG-VVAEGGLDVVTTDSGS  546
           |:      ||:   |  | |:|         |||::|   | |  :||:|  |  |:|
Db   1143  SEKDFTYSLQDTLTGLTSITLLG--GTANGRN---------DTGTVINKDGLTITLANGAA 1191

Qy    547  IGT------KTIGVTGNDVGEGTVHFQNGNGATLGSLYVNV--------TEGNVAFKNFE  592
           ||        |::|:|| || :    :|||: ||   |           |||:|:| ::
Db   1192  AGTDASNGNTISVTKDGISAGNKEITNVKSAL--KTYKDTQNTADETQDKEFHAAVKNAN 1249

Qy    593  LVSKVGQYGQSPDTKLDLNVSTTVEYQ----------LSKYTSDRVYSDPNLEGYEVESK  643
           |    |  |:  |  :|::| : :|:          |: |  |:::    :|   |:|:
Db   1250  EVEFVGKNGATVSAKTDNNGKHTVTIDVAEAKVGDGLEKDTDGKIKLKVDNTDG------N 1304

Qy    644  NLAVADAKIVGNKVVVTGKTPGKVDIHLTKNGATAGKAITEIVQETIAIKSVN--------  696
                      :|: ||:|:        :: |||   | :  |: :::: |
Db   1305  NLLTVDA--------TKGASVAKGEPNAVTTDATTAQGTNANERGKVVVKGSNGATATET 1356

Qy    697  -FKPVQTENFVEKKIN-IGTVLELE---KSNLDD-----IVKGINLTKETQHKVR  741
                | ::| ::|      | :|:|     |::||    :||:| | |: |:| :
Db   1357  DKKKKVATVGDVAKAINDAATFVKVENDDSATIDDSPTDDGANDALKAGDTLTLKAGKNLK 1416

Qy    742  V----------------VKSGAEQGKLYLDRNGDAVFNAGDVK-LGDVTVSQTSDSA---  781
                            |::    |::  :| |||:       | ||: |:|: |:
Db   1417  VKRDGKNITFALANDLSVKSATVSDKLSLGTNGNKVNITSDTKGLNFAKDSKTGDDANIH 1476

Qy    782  LPNFKADLYDTL  793
           |   :| | |||
Db   1477  LNGIASTLTDTL 1488
```

```
RESULT 13
US-10-193-764-61
; Sequence 61, Application US/10193764
; Publication No. US20030113943A1
; GENERAL INFORMATION:
;  APPLICANT: Loosmore, Sheena M.
;  APPLICANT: Yang, Yan-Ping
;  APPLICANT: Klein, Michel H.
;  TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
;  TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
;  FILE REFERENCE: 1038-1239MIS
;  CURRENT APPLICATION NUMBER: US/10/193,764
;  CURRENT FILING DATE: 2002-07-12
;  PRIOR APPLICATION NUMBER: 09/167,568
;  PRIOR FILING DATE: 1998-10-07
;  NUMBER OF SEQ ID NOS: 91
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
;  LENGTH: 1180
;  TYPE: PRT
;  ORGANISM: Haemophilus influenzae
US-10-193-764-61

Query Match             5.9%;  Score 246; DB 14;  Length 1180;
Best Local Similarity  20.2%;  Pred. No. 1.9e-08;
Matches  185; Conservative 124; Mismatches 335; Indels 270; Gaps  41;

Qy     48  VIFTKILNLPVDENAQPSFK------DAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKID  102
           :|:||:                   : ||: :
Db    262  ISFTTDTTFNVNKNAKVNFNIKAPIGTINQYNNLNYALFNGNISVSGGGNVTPRLNASSS  321

Qy    103  RASFASMLVSAYNLKDKVNGELVTTFEDLLDHWGEEKANILINLGISV-GTGGKWEPNKS  161
           :: :      ::        ::    : | : || ::   :|: ||::  ||   ||:
Db    322  NQQTPGVIINSKHLNASKGSSL--RFETT---GSTKVGFLINNDLTLNATGG----NIS  371

Qy    162  VSRAE--------AAQFIALT--DKKYGKKDNAQAYVTDVKVSEPTKLLT-TGTGLDK  208
           ||:           | |:|    ::|:: :: : : | |: :    |:   |||:
Db    372  LLQVEGIDGMIGEGVVAKKNIFTGGNITFGGNITFGSKKAITEIKGNVTINENTNATLIGSDFND  431

Qy    209  LSADDVTLEGDKAVAIEASTDGTSAVVTLGGKVAPNKDLTVKVKNQSFVTKFVEVKKLA  268
               ::   |: :| | |||  :|   :|  :    ::: |:::    :|  :::
Db    432  -HKKPLNIKGD--VVNRGNLTAGGNVINIGGNLTVENGANLKA------ITNFTFNVGGLF  483

Qy    269  VEKLTFDDDRAGQAIAFKLNDEKGNADVEYLNLANHDVKF------VANNLDGSPA  318
           | |   ::: | |    |:|::   |:  :  |:  : :         :|::|
Db    484  -------------------NNKGNSNI---SIARGGAKFKDINNTSSLNITNSDTTYR  520

Qy    319  NIFEGGEATSTTGKLAVGIKQGDYKVEVQV-------TKRGGLTVSNTGI-----ITVK--  365
           |:  |: |:  |:| ||:: ||:    :       |:|||| :|:|  |     :|::
Db    521  TIIB-GNINKAGDLNIIDNKGN--AEIQIGGN-SQKEGNLTISSDKINITKQITIKKGV  577

Qy    366  ---NLDTPASA------------IKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVV-  409
                |:: |:|            :||: :||: :::| | | :::: |: |:|::|::
Db    578  NGENSDSSTKSQANLTIKTKELKLTQD---LNISGFNKAKIVAKDSSNLTIGNSDDSGN  633

Qy    410  -------------GEKASLNKLVATIAGEDKVVDPGS----------  433
                        |:||:|: | |:  | |:|
Db    634  TSAKTVTFNNVKDSKISADGHKVTLNSKVKTLSDNDNNTEGGSDNNTGLTITAKDVEVNN  693

Qy    434  ------ISIKSSNHGIISVVNNYITAEAAGEATLTIKVGDVTKDVKFPK---  475
                 :::::| |  ||:| : : ||||  ||||::| |::  : |: |:
Db    694  NITSHKTVNVSAANGGITTKTGTTINA-TAGNVEITAHTGSIQGGIESKPGSVTIVAGGD  752

Qy    476  ---VTTDSRKLVSVKANPDKLQVVQNKTLPVT--FVTTDQYGDPFG---ANTAAIK---  523
              |:|:  |  |  |||: :: ::|:  |     |::|  || |      | ||:
Db    753  TLAVGNISGNAVTVTANSGALTTLAGSTIKGTESITTSSQSGNIGGKISGKTVNVKATNS  812

Qy    524  --------------EVLPKT-----GVVAEGGLDV----------VTTDSGSIGTKTIGVTGND  558
                         :|||:     | |:|| ||:          |:|:| |:|:| |::|
Db    813  LTTQADSKIEATEGBANVTSKTSIIGGTISGGTVEVTATEGLTTQAGSTITGTESVTTSS  872
```

```
Qy   559 VGEGTVHFQNGN-GATL--GSLYVNVTEGNVAFKNFELVSKVGQYGQSPDTKLDLNVSTT 615
             :|||| ::|:||: |:    ::|||::|  |    ||::|   :
Db   873 ------QSGNIGGMISGGKVEVSATK-------DLIITKSG--SEIKATAGEVNVTSA 914

Qy   616 VEYQLSKYTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVDIHLTKNG 675
Db   915 T------------------------GTIDGTISGNTVNVTANTGDLTVEDAAKID 945

Qy   676 ATAGKATVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLELEKSNLDDIVKGINLTKE 735
             :|||:|  :||::|:|    :: |  ::|
Db   946 ATGGAATLTATSGKLTTKASS--SITSAN---NQVN----LSAKDGSIGGNINAANVTLN 996

Qy   736 TQHKVRVVKSGAEQGKLYLDRNGDAVFNAGDVKL-----GDVTVSQTSDSALPNFKADLY 790
             ::: |||      |  ||  |:    || |:    ||:|
Db   997 TTGALTTVK----GSSINANSGTLVINAKDAELNGEASGNHTVVNATNA---NGSGSVI 1048

Qy   791 DLITKYTDKGTLV 804
             :|||
Db  1049 ATTSSRVNITGDLI 1062
```

RESULT 14
US-10-193-764-59
; Sequence 59, Application US/10193764
; Publication No. US20030133943A1
; GENERAL INFORMATION:
;   APPLICANT: Loosmore, Sheena M.
;   APPLICANT: Yang, Yan-Ping
;   APPLICANT: Klein, Michel H.
;   TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
;   TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
;   FILE REFERENCE: 1038-1239MIS
;   CURRENT APPLICATION NUMBER: US/10/193,764
;   CURRENT FILING DATE: 2002-07-12
;   PRIOR APPLICATION NUMBER: 09/167,568
;   PRIOR FILING DATE: 1998-10-07
;   NUMBER OF SEQ ID NOS: 91
;   SEQ ID NO 59
;   LENGTH: 1188
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-10-193-764-59

Query Match   5.9%;   Score 246;   DB 14;   Length 1188;
Best Local Similarity 20.2%;   Pred. No. 1.9e-08;
Matches 185;   Conservative 124;   Mismatches 335;   Indels 270;   Gaps 41;

```
Qy    48 VIFTKILNLPVDENAQPSFK-----DAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKID 102
             :  :|:|  |
Db   270 ISFTTDTTFNVNKQNAKVNFNIKAPIGTINQYNNLNYALFNGNISVSGGGNVTFRLNASSS 329

Qy   103 RASFASMLVSAYNLKDKVNGELVTTFEDLLDHWGEEKANILINLGISV-GTGGKWEPNKS 161
             :::|     |||    ||| |:
Db   330 NQQTPGVIINSKHLNASKGSSL--RFETT----GSTKVGFLINNDLTLNATGG----NIS 379

Qy   162 VSRAE-------------AAQFIALT--DKKYGKKDNAQAYVTDVKVSEPTKLTLTGTGLDK 208
             : :|
Db   380 LLQVEGIDGMIGEGVVAKKNITFTGGNITFGSKKAITEIKGNVTINENTNATLIGSDFND 439

Qy   209 LSADDVTLEGDKAVAIEASTDGTSAVVTLGGKVAPNKDLTVKVKNQSFVTKFVYEVKKLA 268
             : :||            |: ||
Db   440 -HKKPLNIKGD--VVNRGNLTAGGNVINIGGNLTVENGANLKA-----ITNFTNVGGLF 491

Qy   269 VEKLTFDDDRAGQAIAFKLNDEKGNADVEYLNLANHDVKF----------VANNLDGSPA 318
             ::::| :                                          |||:
Db   492 -------NNKGNSNI---SIARGGAKFKDINNTSSLNITTNSDTTYR 528

Qy   319 NIFEGGEATSTTGKLAVGIKQGDYKVEVQV-----TKRGGLTVSNTGI--ITVK---- 365
             :::: |  :::   |:: | :||  ||  |   :|||  |::|||:|    ||:|
Db   529 TIIE-GNITNKAGDLNIIDNKGN--AFIQIGGNISQXEGNLTISSDKINITKQITTIKKGV 585

Qy   366 ---NLDTPASA------IKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVV-------- 409
```

```
Db   586 NGENSDSSTKSQANLTIKTKELKLTQD----LNISGFNKAKIVAKDSNLTIGNSDDSGN 641
             :|:   :                      |::|||:|:
Qy   410 ------------------------GEKASLNKLVATIAGEDKVVDPGS------------ 433

Db   642 TSAKTVTFNVKDSKISADGHKVTLNSKVKTLSDNDNNTEGGSDNNTGLTITAKDVEVNN 701
             :::|                 :: ||:|::
Qy   434 ------ISIKSSNHGIISVVNNYITAEAAGEATLTIKVGDVTKDVKFK--------- 475

Db   702 NITSHKTVNVSAANGGITTKTGTTINA-TAGNVEITAHTGSIQGGIESKPGSVTIVAGGD 760
Qy   476 ---VTTDSRKLVSVKANPDKLQVVQNKTLPVT--FVTTDQYGDPFG--ANTAAIK---- 523

Db   761 TLAVGNISGNAVTVTANSGALTTLAGSTIKGTESITTSSQSGNIGGKISGKTVNVKATNS 820
Qy   524 ------------EVLPKT----GVVAEGLDV----VTTDSGSIGTKTIGVTGND 558

Db   821 LTTQADSKIEATEGEANVTSKTSIIGGTISGGTVEVTATEGLTTQAGSTITGTESVTTSS 880
             :|||
Qy   559 VGEGTVHFQNGN-GATL--GSLYVNVTEGNVAFKNFELVSKVGQYGQSPDTKLDLNVSTT 615

Db   881 ------QSGNIGGMISGGKVEVSATK-------DLIITKSG--SEIKATAGEVNVTSA 922
Qy   616 VEYQLSKYTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVDIHLTKNG 675

Db   923 T------------------------GTIDGTISGNTVNVTANTGDLTVEDAAKID 953
Qy   676 ATAGKATVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLELEKSNLDDIVKGINLTKE 735

Db   954 ATGGAATLTATSGKLTTKASS--SITSAN---NQVN----LSAKDGSIGGNINAANVTLN 1004
Qy   736 TQHKVRVVKSGAEQGKLYLDRNGDAVFNAGDVKL-----GDVTVSQTSDSALPNFKADLY 790

Db  1005 TTGALTTVK------GSSINANSGTLVINAKDAELNGEASGNHTVVNATNA---NGSGSVI 1056
Qy   791 DLITKYTDKGTLV 804

Db  1057 ATTSSRVNITGDLI 1070
```

RESULT 15
US-10-193-764-28
; Sequence 28, Application US/10193764
; Publication No. US20030133943A1
; GENERAL INFORMATION:
;   APPLICANT: Loosmore, Sheena M.
;   APPLICANT: Yang, Yan-Ping
;   APPLICANT: Klein, Michel H.
;   TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
;   TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
;   FILE REFERENCE: 1038-1239MIS
;   CURRENT APPLICATION NUMBER: US/10/193,764
;   CURRENT FILING DATE: 2002-07-12
;   PRIOR APPLICATION NUMBER: 09/167,568
;   PRIOR FILING DATE: 1998-10-07
;   NUMBER OF SEQ ID NOS: 91
;   SEQ ID NO 28
;   LENGTH: 1220
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-10-193-764-28

Query Match   5.8%;   Score 242.5;   DB 14;   Length 1220;
Best Local Similarity 21.5%;   Pred. No. 3.6e-08;
Matches 210;   Conservative 139;   Mismatches 368;   Indels 259;   Gaps 52;

```
Qy     2 GKSFPDVPAGHWAEGS-INVLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE 60
             |::|     |:|:||  :|:|        :||||:|          :|:
Db   218 GRTYWNVTHLNVSEGSKFNLTID----SRGDDTAG------------TLNTPYNL 256

Qy    61 NAQPSFKDAKNIWSSKYIAAVE----KAGV------VKGDGKENFYPEG 99
```

```
Db  257 NGISFNKD--TIFDVKQNGAVTFDIKAPIGVNNRNLNYASFNGNISVSGGGNVTFKLLA 314
             :   :::  : ::   :  ::   : :::: ::  ::  : :
Qy  100 KIDRASFASMLVSAYNLKDKVNGELVTTFEDLLDHWGEEKANILINLGISV-GTGGKWEP 158
                                                           |||  |||
Db  315 SSSTAQTPGVFINSKHFNASGGSSLEFRTE------GSTKVGFLINNDLTLNATCG---- 364
               :  ::: :                          :::
Qy  159 NKSVSRAEAAQFI---ALTDKK---------YGKKDNAQAYVTDVKVSEPTKLTLTGTG 205
        |  |:  :                        :|  :||        ::||:|:||||:
Db  365 NISLLQVEGIDGMIGKGVVAKKNITFAGGNITFGSKKAITEIEGNATINNNANVTLIGSD 424
                                    : :||| :|:   |  ::
Qy  206 LDKLSADDVTLEGDKAVAIEASTDGTSAVVTLGGKVAP-NKDLTV-KVKNQSFVTKFVYE 263
                                     :::|:  :||: |  : ||:|
Db  425 FD--------NHQKPLTIKKDVIINSGNLTAGGNVININGNLTVNNGANLKAITNFTFN 475
                                                        ||
Qy  264 VKKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLNLANHDVKF---------VANNL 313
          ||:::   ::|                       ::    :::
Db  476 VGGLF--------------DNKGNSNI---SIARGGAKFKDINNTSSLNITTNS 512
                                                        :::
Qy  314 DGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQV-----TKRGGLTVSN-----TGIII 363
                                            :::       :|::|:|   |||::
Db  513 DTTYRTIIE-GNITNKAGDLNIIDNKGN--AEIQIGGNISQEGNLTISSDKINITNQIT 569
                                                          :  :||::
Qy  364 VK----NLDTPASAIKNVFAL--DADNDGVVNYGSKLSGKD---FALNSQNLVVGEK 412
          :                       : |    :|:         :::|||:||||
Db  570 IKKGVNKEDSDSSTANNANLTIKTKELQLTGDLN---ISGPDKABITAKEGADLIIGNS 625
                  :  :: ::  : :|   :                  :|||:  ||::||:
Qy  413 ASLN----KLVATIAGEDKVVDPGSISIKSSNHGIISVVNNYITAEAAGEATLTIKVGDV 468
                                   : ::  :   :        :::||    : :
Db  626 DNNNNANAKKVTFNQVKDSKISAGSHNVTLNSKVETSNGNNDAESNNGDSTSLTINAKNV 685
                                                  ::  :||: :::::||:
Qy  469 TKDVKFKVTTDSRKLVSVKANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPK 528
                                              ::  |::|:
Db  686 T--VNNNIT--SHKTVNITASEN------------VTTKAGTTNATIGSV-EVTAK 725
                                                    ::|:|:|||: |:|:|
Qy  529 TGVVAEGLDVVTTDSGSIGTKTIG-------VTGNDVGEGTVHFQNGNGATLGSLYVNV 581
        | |:::|:|  : ::||:           ||  |||  ||||  |||| ::| :::
Db  726 TGDI-KGGIE---SNSGNVNITASGDTLNVSNITGQNV---TVAAASGAVTTTKGSTINA 778
                                            |:                   :|:
Qy  582 TEGNVAFKNFELVSKVGQY-GQSPD'TKLDLN-------------------VST 614
        ||                      ||    ::                     |||
Db  779 TTGNA-----NTTKTGEINGEVKSASGNVNITASGNTLNVSNITGQNVTVTANSGAITT 833
                                                                 :|:|:|
Qy  615 TVEYQLSKYTSD-RVVSDPENLEGYEVESKNLAV-------------ADAKIVGNKVVVT---G 661
        |  |:::   : ::|:|:                             : |:||||:|||   |
Db  834 TEGSTINATTGDANITTQTGNING-KVESSSGSVTLIATQTLAVGNISGDTVIITADKG 892
                                                             :      :|:
Qy  662 K--TPGKVDIHLTKNGAT--AGKATVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVL 716
        |                  :::|   :    : |:|    : |||||:|||: : : |::|
Db  893 KLTTQTSSKINGTKSVTTSSCQSGDISGTISGNTVSVSATGSLTTQAGSKIEAKTGEANVT 952
                                                       :::|:||:|| |::||:
Qy  717 ELEKSNLDDIVKG--INLTKETQHKVRVVKSGAE-------------QGKLYLDRNGDAV 761
         ::::::::  |                                      ::::  |||
Db  953 S-ATGTIGGTISGNTVNVTANTDN--LTIKDGARIKATGGAVTLTATGGTLTTETSSDIT 1009
                                                    ||||||  ::||::|:::||
Qy  762 FNAGDVKL--GDVTVSQTSDSALPNFKADLYDTLTTKYTDK-----GTLVFKVLKDKDVI 814
         |   :::     ::|||  : :::: |||: ::|||::::     |||::::: :|::|
Db 1010 SSNGQTTLTAKDSSIAGSINAA--NVTLNTTGTLTTVAGSKIEAASGTLVINA-KDAQLD 1066
                 :  :::: : ::::|:

Qy  815 TSEIGSQAVHVNVLN 829
             |||:||: |:|:|||
Db 1067 GAALGDR-TEVNVTN 1080
```

Search completed: April 7, 2004, 17:30:49
Job time : 50 secs

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2004, 17:30:55 ; Search time 23 Seconds
                                     (without alignments)
                                     1869.758 Million cell updates/sec

Title: US-09-844-281-1
Perfect score: 4202
Sequence: 1 AGKSFPDVPAGHWAEGSINY..........ITSEIGSQAVHVNVLNNPNL 833

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_AA:*
               1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
               2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
               3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
               4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
               5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
               6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	374	8.9	1222	2	US-08-682-517-15	Sequence 15, Appl
2	374	8.9	1252	2	US-08-682-517-9	Sequence 9, Appli
3	367	8.7	920	4	US-09-463-402-6	Sequence 6, Appli
4	362.5	8.6	921	4	US-09-889-572-4	Sequence 4, Appli
5	257.5	6.1	1338	2	US-08-728-470-9	Sequence 9, Appli
6	257.5	6.1	1338	2	US-08-719-641-9	Sequence 9, Appli
7	257	6.1	1599	2	US-08-617-697-9	Sequence 9, Appli
8	251.5	6.0	1529	2	US-08-728-470-10	Sequence 10, Appl
9	251.5	6.0	1529	3	US-08-719-641-10	Sequence 10, Appl
10	249.5	5.9	1600	2	US-08-617-697-10	Sequence 10, Appl
11	249.5	5.9	2353	3	US-09-377-155-33	Sequence 33, Appl
12	249.5	5.9	2353	3	US-08-913-942-4	Sequence 4, Appli
13	249.5	5.9	2353	4	US-09-669-974-33	Sequence 33, Appl
14	249.5	5.9	2353	4	US-09-797-862-33	Sequence 33, Appl
15	247.5	5.9	2354	4	US-09-268-347-47	Sequence 47, Appl
16	246	5.9	1180	4	US-09-206-942-65	Sequence 65, Appl
17	246	5.8	1188	4	US-09-206-942-63	Sequence 63, Appl
18	244.5	5.8	1912	1	US-08-400-995-4	Sequence 4, Appli
19	244.5	5.8	1912	3	US-08-685-467-4	Sequence 4, Appli
20	242.5	5.8	1220	4	US-09-206-942-28	Sequence 28, Appl
21	242.5	5.8	1226	4	US-09-206-942-26	Sequence 26, Appl
22	240.5	5.7	2411	4	US-09-268-347-36	Sequence 36, Appl
23	237.5	5.7	2314	4	US-09-268-347-49	Sequence 49, Appl
24	235.5	5.6	2048	3	US-09-268-347-48	Sequence 48, Appl
25	235	5.6	1004	4	US-09-206-942-57	Sequence 57, Appl
26	235	5.6	1010	4	US-09-206-942-55	Sequence 55, Appl
27	229	5.4	969	4	US-09-206-942-32	Sequence 32, Appl
28	229	5.4	975	4	US-09-206-942-30	Sequence 30, Appl
29	227.5	5.4	1073	4	US-09-206-942-49	Sequence 49, Appl
30	227.5	5.4	1079	4	US-09-206-942-47	Sequence 47, Appl
31	227	5.4	2504	4	US-09-328-352-5821	Sequence 5821, Ap
32	222	5.3	1833	4	US-08-621-944A-4	Sequence 4, Appli
33	222	5.3	1833	4	US-08-945-567D-4	Sequence 4, Appli
34	222	5.3	1992	4	US-08-621-944A-3	Sequence 3, Appli
35	222	5.3	1992	4	US-08-945-567D-3	Sequence 3, Appli
36	219.5	5.2	2315	4	US-09-543-681A-5434	Sequence 5434, Ap
37	217.5	5.2	1005	4	US-09-206-942-41	Sequence 41, Appl
38	217.5	5.2	1011	4	US-09-206-942-39	Sequence 39, Appl
39	216.5	5.2	1095	4	US-09-206-942-69	Sequence 69, Appl
40	216.5	5.2	1536	1	US-08-038-682-2	Sequence 2, Appli
41	216.5	5.2	1536	2	US-08-302-832-2	Sequence 2, Appli
42	216.5	5.2	1536	2	US-08-530-198-2	Sequence 2, Appli
43	216.5	5.2	1536	2	US-08-469-880-2	Sequence 2, Appli
44	216.5	5.2	1536	2	US-08-728-470-2	Sequence 2, Appli
45	216.5	5.2	1536	2	US-08-617-697-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-682-517-15
; Sequence 15, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; -APPLICANT:
;   TITLE OF INVENTION: Expression of surface layer proteins
;   NUMBER OF SEQUENCES: 25
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/682,517
;   FILING DATE:
;   CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 15:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1222 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-682-517-15

Query Match   8.9%; Score 374; DB 2; Length 1222;
Best Local Similarity   23.3%; Pred. No. 2.7e-19;
Matches 224; Conservative 130; Mismatches 345; Indels 264; Gaps 47;

Qy   13 WAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQPSFKDAK-N 71
            ||:|| | |: :|| :| | : ||: | ::
Db   12 YAKEAVQALVDQGVIQGDTNGFNPLNTVTRAQAAEIFTKALEL--EANGDVNFKDVKAG 69

Qy   72 IWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASPASMLVSAYNLKDKVNGELVTTFEDL 131
            || | :|| : : | | : | || || | |||
Db   70 AWYYNSIAAVVANGIFEGVSATEFAPNKSLTRSEAAKILVEAFGLEGEAD---LSEFADA 126

Qy   132 --LDHWGEEKANILINLGISVGT-GGKWEPNKSVSRAE-AAQFIALTDKKYGKKDNAQAY 187
           | : ||| |:| |: | || | ::| ::
Db   127 SQVKPWAKKYLEIAVANGIFEGTDANKLNPNNSITRQDFALVFKRTVDKVEGETPEEAAF 186

Qy   188 V------TDVKVS---EPTKLTLTGTGLDKLSADDVTL-EGDKAVAI---EASTDGTSAVV 235
                      | | | | | ||
Db   187 VKAINNTTVEVTFEEEVTNVQALNFKIEGLEIKNASVKQTNKKVVVLTTEAQTADKEYVL -246

Qy   236 TLGGK---------VAPNKDLTVKVKNQSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAF 285
             ||:| ||| | :: || |::|
Db   247 TLDGETIGGFKGVAAVVPTK---VELVSSAVQGKLGQEVKVQAKVTVAEGQSKAGIPVTF 303

Qy   286 -----------KLNDEKGNADVEYINLANHDVKFVA----NNLDGSPANIFEGG 324

```
Query Match      8.9%;  Score 374;  DB 2;  Length 1252;
Best Local Similarity  23.3%;  Pred. No. 2.8e-19;
Matches 224;  Conservative 130;  Mismatches 345;  Indels 264;  Gaps  47;

Qy    13 WAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQPSFKDAK-N  71
         ::|::.:||:|:  :  . |||.|     :   ||::|||::|:: : ||:|:  :
Db    42 YAKEAVQALVDQGVIQGDTNGNFNPLNTVTRAQAAEIFTKALEL--EANGDVNFKDVKAG  99

Qy    72 IWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGELVTTFEDL 131
         |||: : :| . |   |.   :  ||||      :   ||::  :     ::SEFADA
Db   100 AWYYNSIAAVVANGIFEGVSATEFAPNKSLTRSEAAKILVEAFGLEGEAD---LSEFADA 156

Qy   132 --LDHWGEEKANILLNLGISVGT-GGKWEPNKSVSRAE-AAQFIALTDKKYGKKDNAQAY 187
           :: |     :|||.| :     . || || |:|  .|:|||: :|::|: :
Db   157 SQVKPWAKKYLEIAVANGIFEGTDANKLNFNNSITRQDFALVFKKRTVDKVEGETPEEAAF 216

Qy   188 V------TDVRVS---EPTKLTLTGTGLDKLSADDVTL-EGDKAVAI---EASTDGTSAVV 235
         :       |:| |   ||: |:||||:| |:|:| |:| |:|.|:|     :  .|
Db   217 VKAINNTTVEVTFEEEVTNVQALNFKIEGLEIKNASVKQTNKKVVVLTTEAQTADKEYVL 276

Qy   236 TLGGK-------VAPNKDLTVKVKNQSFVTKPVYEVKKLAVEKLTFDDDRAGQAIAF 285
         ||||         :  || |.  . :||: |||  |  . :  :  |
Db   277 TLDGEFIGGFKGVAAVVPTK---VELVSSAVQGKLGQEVKVQAKVTVAEGQSKAGIPVTF 333

Qy   286 ---------------KLNDEKGNADVEYLNLANHDVKFVA------NNLDGSPANIFEGG 324
                          |:|  | | | :|  . | ||        ::|||    ||
Db   334 TVPGNNNDGVVPTLTGEALTNEEGIATYSYTRYKEGTDEVTAYATGDRSKFSLGYVFWGV 393

Qy   325 EATSTTGKLAVG--IKQG---DYKVEVQVTKRGGLTVS---NTGIITVKNL--DTPASAI 374
         ||:|||| |: |  |:|    ||||:| |:||:  |||    :||   :|   :|:||:|
Db   394 DTILSVEEVTTGASVNNGANKTYKVTYKNPKTGKPEANKTFNVGFVENMNVTSDKVANAT 453

Qy   375 KNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLNKLVATIAGEDKVVDPGSI 434
         :  |||| | |  :   |  |    |  :   : | | :   | |   |  ||||:
Db   454 VNGVKALQLSN------GTALDAAQITTDSK---------GEATFTVSGTNAAVTPVVY 497

Qy   435 SIKSSNHGIISVVNNYITAE--------AAGEATLTIKV------GDV-----TKDV 472
         ||:||:: |: | | ||: |          :|::  |  |      |        ||.
Db   498 DLHSTNN--STSNKKYSASALQTTASKVTFAALQAEYTIELTRADNAGEVAAIGATNGR 554

Qy   473 KFKVTTDSR-------KLVSVKANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEV 525
         |||:||||        |:| | |  |:  ||:  :|  |  |  . :  :  |
Db   555 EYKVIVKDKAGNLAKNEIVNVAFNEDKDRVISTVT-NAKFVDTD------------- 597

Qy   526 LPKTGVVAEGG---------LDVVTTDSGS----IGTKT---------IGVTGNDVGEGTV--- 564
         | ||.| |||           |||: |::      |.|            :|. ||| ||:
Db   598 -PDTAVYFTGDKAKQISVKTNDKGEATFVIGSDTVNDYATPIAWIDINTSDAKQGDLDEG 656

Qy   565 ----------HFQ------NGNG------ATLGSLYVNVTEGNVAFKNFELVSKVGQ----- 599
                   |:|      ||||        |.||: |    || |:: |  |:|
Db   657 EPKAVAPISYFQAPYLDGSAIKAYKKSDLNKAVTKFDGSETAVFAAELVNQSGKKVTGTS 716

Qy   600 ---------YGQ-SPDTKLDLNV-----STTVEYQLSKYTSDRVYSDPENLEGYEVESKNL 645
                  ||| |||| || |     |  || .|::      | .| |:    |  |
Db   717 IKKATYIYNTGANDIKVDNQVISPNRSYTVTYEATLSSTGTVITPAKNLBVTSVDGKTT 776

Qy   646 AVADAKIVGNKVVVTG---KTPGKVDIHLTKNGATAG-KATVEIVQETIAIKSVNFKPVQ 701
         || ||:||||| | |    |||||:|| |:||| |   |||.|:|| |   |::|| ||
Db   777 AV------KVIATGIAVNTDGK-DYAFTAKEATATFTATNEVPN-------SYTGVA 819

Qy   702 TE-NFVEKKINIGTVLELEKSNLDDIVKGINLTKETQHKVRVVKSGAEQGKL--YLDRNG 758
          | || |||:|||| | ::||| | || || ||: || |:|| |  |:|      ||| |
Db   820 TQFNTADSGSNSNIWFAGKN---------------------------PVKYAGVSGKTYKYFGANG 859

Qy   759 DAVFNAGDVKLGDVTVSQTSDSALPNFKADLYDTLTTKYTDKG---TLVFKVLKDKDVIT 815
         |||| |||||||||| ||||||||||||||||| | |:||:|      |.|:|||:||:|
Db   860 NEVFGEA------------------------------AWEALLTQYATEGQKVTISYNV--DGDTVT 894

Qy   816 SEI 818
         ::
Db   865 FKV 867
```

Lower alignment block (duplicate / continuation):

```
Db   304 TVPGNNNDGVVPTLTGEALTNEEGIATYSYTRYKEGTDEVTAYATGDRSKFSLGYVFWGV 363
         :|:|:||
Qy   325 EATSTTGKLAVG--IKQG---DYKVEVQVTKRGGLTVS---NTGIITVKNL--DTPASAI 374
         ||:|||| |: |  |:|    ||||:| |:||:  |||    :||   :|   :|:||:|
Db   364 DTILSVEEVTTGASVNNGANKTYKVTYKNPKTGKPEANKTFNVGFVENMNVTSDKVANAT 423

Qy   375 KNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLNKLVATIAGEDKVVDPGSI 434
         :  |||| | |  :   |  |    |  :   : | | :   | |   |  ||||:
Db   424 VNGVKALQLSN------GTALDAAQITTDSK---------GEATFTVSGTNAAVTPVVY 467

Qy   435 SIKSSNHGIISVVNNYITAE--------AAGEATLTIKV------GDV-----TKDV 472
         ||:||:: |: | | ||: |          :|::  |  |      |        ||.
Db   468 DLHSTNN--STSNKKYSASALQTTASKVTFAALQAEYTIELTRADNAGEVAAIGATNGR 524

Qy   473 KFKVTTDSR-------KLVSVKANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEV 525
         |||:||||        |:| | |  |:  ||:  :|  |  |  . :  :  |
Db   525 EYKVIVKDKAGNLAKNEIVNVAFNEDKDRVISTVT-NAKFVDTD------------- 567

Qy   526 LPKTGVVAEGG---------LDVVTTDSGS----IGTKT 564
         | ||.| |||           |||: |::      |.|
Db   568 -PDTAVYFTGDKAKQISVKTNDKGEATFVIGSDTVNDYATPIAWIDINTSDAKQGDLDEG 626

Qy   565 ----------HFQ------NGNG------ATLGSLYVNVTEGNVAFKNFELVSKVGQ----- 599
                   |:|      ||||        |.||: |    || |:: |  |:|
Db   627 EPKAVAPISYFQAPYLDGSAIKAYKKSDLNKAVTKFDGSETAVFAAELVNQSGKKVTGTS 686

Qy   600 ---------YGQ-SPDTKLDLNV 645
                  ||| |||| || |
Db   687 IKKATYIYNTGANDIKVDNQVISPNRSYTVTYEATLSSTGTVITPAKNLBVTSVDGKTT 746

Qy   646 AVADAKIVGNKVVVTG---KTPGKVDIHLTKNGATAG-KATVEIVQETIAIKSVNFKPVQ 701
Db   747 AV------KVIATGIAVNTDGK-DYAFTAKEATATFTATNEVPN-------SYTGVA 789

Qy   702 TE-NFVEKKINIGTVLELEKSNLDDIVKGINLTKETQHKVRVVKSGAEQGKL--YLDRNG 758
Db   790 TQFNTADSGSNSNIWFAGKN---------------------------PVKYAGVSGKTYKYFGANG 829

Qy   759 DAVFNAGDVKLGDVTVSQTSDSALPNFKADLYDTLTTKYTDKG---TLVFKVLKDKDVIT 815
Db   830 NEVFGEA------------------------------AWEALLTQYATEGQKVTISYNV--DGDTVT 864

Qy   816 SEI 818
         ::
Db   865 FKV 867
```

```
RESULT 3
US-09-463-402-6
; Sequence 6, Application US/09463402
; Patent No. 6596510
; GENERAL INFORMATION:
; APPLICANT: Lubitz, Werner
; APPLICANT: Resch, Stephanie
; TITLE OF INVENTION: Secretion of Carrier-bound Proteins into the Periplasm
; TITLE OF INVENTION: and into the Extracellular Space
; FILE REFERENCE: 05649059
; CURRENT APPLICATION NUMBER: US/09/463,402
; CURRENT FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: DE19732829.6
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: PCT/EP98/04723
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 6
; LENGTH: 920
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-09-463-402-6

  Query Match             8.7%;  Score 367;  DB 4;  Length 920;
  Best Local Similarity  23.3%;  Pred. No. 5.8e-19;
  Matches  227;  Conservative 138;  Mismatches 369;  Indels 240;  Gaps  48;

Qy      4 SFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQ 63
            ||:||   :::|:|| ||
Db     33 SFTDVAPQY--KDAIDFLVSTGATKGKTETKFGVYDEITRLDAAVILARVLKLDVDNAKD 90

Qy     64 PSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGE 123
            ::|:|:|    ||||   :|:||    |:
Db     91 AGFTDVPKD-RAKYVNALVEAGVLNGKAPGKFGAYDPLTRVEMAKIIANRYKLK---ADD 146

Qy    124 LVTTFEDLLDHWGEEKANILINLGISVG-TGGKWEPNKSVSRAPAAQFIALTDKKYGKKD 182
            :|:| |             ::|||    :|
Db    147 VKLPFTDVNDTWA-PYVKALYKYEVTKGKTPTSFGAYQNITRGDFAQFVY------RAV 198

Qy    183 NAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVT-LEGDKAVAIEAST--DGTSAVVTJG 238
            |:|:|                 :::::|    :|:
Db    199 NINAVPEIVEVTAVNSTTVKVTFNTQIADVDFTNFAIDNGLTVTKATLSRDKKSVEVVVN 258

Qy    239 GKVAPNKDLTVK---VKN-----QSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAFKLND 289
            |:|||||        :||      :
Db    259 KPFTRNQEYTITATGIKNLKGETAKELTGKFVWSVQDAVTVALNNSSLKVGEESGLTVKD 318

Qy    290 EKGNADVEYLNLANHDVKFVANNLDGSPANI--FEGGEATSTTGKLAVGIKQGDYKVEVQ 347
            ::|:|       ::||:|     | ::
Db    319 QDGK--------DVVGAKVELTSSNTNIVVSSGEVSVSAAKV-TAVKPGTADVTAK 366

Qy    348 VTKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNL 407
            :::|| |:::|:|    |
Db    367 VTLPDGVLTNFKVTVTEVPV--QVQNQGFTL-VDN-------LSNA----PQNT 408

Qy    408 VVGEKASLNKLVATIAGEDKVV---------DPGS------ISIKSSNHGIISVNN---- 449
            |   :|   :|:
Db    409 VAFNKA--EKVTSMFAGETKIVAMYDTKNGDPETKPVDFKDATVRSLNPIIATAAINGSE 466

Qy    450 -YITAEA--AGEATLTIKVGDVTK------DVK-------PKVTTDSRKLVSVKANPDKLQ 494
             ::|:||                         |
Db    467 LLVTANAGQSGKASFEVTFKDNTKRFTVDVKKDPVLQDIKVDATSVKLSDEAVGGGEVE 526

Qy    495 VVQNKLPLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVABGGLDVVTTDSGSIGTKTIGV 554
            ::   |::|:
Db    527 GVNQKTIKVSAV--DQYG---------KEI--KFGT--KGKVTVTTNTEGLVIKNVN- 568

Qy    555 TGNDVGEGTVHFQNGNGATLGSLYV----NVTEGNVAFKNFELVS----------- 595
            :|             ::
Db    569 ----SDNTIDFDFGSGNSATDQFVVATKDKIVNGKVKYFKNASDTTPSTKITTVNVV 623

Qy    596 ------KVGQYQGSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEGYEVES----K 643
```

```
RESULT 4
US-09-889-572-4
; Sequence 4, Application US/09889572
; Patent No. 6610517
; GENERAL INFORMATION:
; APPLICANT: Lubitz, Werner
; TITLE OF INVENTION: COMPARTMENTALIZATION OF RECOMBINANT POLYPEPTIDES IN HOST CELLS
; FILE REFERENCE: 100564-00070
; CURRENT APPLICATION NUMBER: US/09/889,572
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: PCT/EP00/00686
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: DE 199 03 345.5
; PRIOR FILING DATE: 1999-01-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
;
; SEQ ID NO 4
; LENGTH: 921
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-09-889-572-4

  Query Match             8.6%;  Score 362.5;  DB 4;  Length 921;
  Best Local Similarity  23.3%;  Pred. No. 1.3e-18;
  Matches  229;  Conservative 135;  Mismatches 364;  Indels 253;  Gaps  48;

Qy      4 SFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDENAQ 63
            ||:||   :::|:|| ||
Db     33 SFTDVAPQY--KDAIDFLVSTGATKGKTETKFGVYDEITRLDAAVILARVLKLDVDNAKD 90

Qy     64 PSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKVNGE 123
            ::|:|:|    ||||   :|:||    |:
Db     91 AGFTDVPKD-RAKYVNALVEAGVLNGKAPGKFGAYDPLTRVEMAKIIANRYKLK---ADD 146

Qy    124 LVTTFEDLLDHWGE--------EKANILNLGISVGTGTGKWEPNKSVSRAEAAQFIALTD 175
            :|:| |            :|
Db    147 VKLPFTDVNDTWAPYVKALYKYEVTKGKLRLKHQQASVHT------KNITLRDFAQFVY-- 196

Qy    176 KKYGKKDNQAYVTDVKVSEPTKLTLTGTGLDKLSADDVT-LEGDKAVAIEAST--DGT 231
            ::
Db    197 ----RAVNINAVPEIVEVTAVNSTTVKVTFNTQIADVDFTNFAIDNGLTVTKATLSRDKK 252

Qy    232 SAVVTLGGKVAPNKDLTVK---VKN------QSFVTKFVYEVKKLAVEKLTFDDDRAGQA 282
            :::
Db    253 SVEVVVNKPFTRNQEYTITATGIKNLKGETAKELTGKFVWSVQDAVTVALNNSSLKVGEE 312

Qy    283 IAFKLNDEKGNADVEYLNLANHDVKFVANNLDGSPANI--FEGGEATSTTGKLAVGIKQG 340
            ::
Db    313 SGLTVKDQDGK--------DVVGAKVELTSSNTNIVVSSGEVSVSAAKV-TAVKPG 360
```

```
Db    624 NVKADATPVGLDIVAP-SEIDVNAPNTAS-------TADVDFINFESVEIYTLDSNGNRLJK 676
                 ::|:|::|:     |:::|::|
Qy    644 NLAVADAKIVGNK--VVVTGK--------------TPGKVDIHLTKNGAT------AGK 680
                |:|:    :::       :|: |
Db    677 KVTPTATTLVGTNDYVEVNGNVLQFKGNDELTLLTSSSTVNVDVTADGTKRIPVKYINS 736
                |    :  |:|:: ::|
Qy    681 ATVEIVQETIAIKSVNFKPVQTEN---FVEKKINIGTVLELEKSNLDDIVKGINLTKETQ 737
                |:|:|      ||||||
Db    737 ASVP-ASATVATSPVTVKLNSSDNDLTFEELIPGVIDPTQLVKD--EDINEFIAVSKAAK 793
                   :|:||||
Qy    738 H-------------------KVRVVKSGAE--------QGKLYLDRN----GDAVFN-- 763
                :     ||          ::
Db    794 NDGYLYNKPLVTVKDASGKVIPTGANVVGLNHDATNGNIWPDEEQAGLAKKFSDVHFDVD 853
                ||:|      :|
Qy    764 --AGDVKLJGDVTVSQTSDSALPNFKADLYDTLTTKYTDKG-----TLVFK-VLKDKDV 813
                   :|            ::| |::
Db    854 FSLANVVKTGSGTVSSS--------------PSLSDAI-QLTNSGDAVSFTLVIKSIYVKGADK 902
                :::  :|
Qy    814 ITSEIGSQAVHVNV 827
                  |:|:||
Db    903 DDNNLLAAPVSVNV 916
```

APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1338 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-728-470-9

Query Match   6.1%;   Score 257.5;   DB 2;   Length 1338;
Best Local Similarity 22.0%; Pred. No. 1.8e-10;
Matches 183; Conservative 104; Mismatches 328; Indels 215; Gaps 38;

```
Qy   16 GSI-----NYLVDKGA---------------ITGKPDGTVGPTESIDRASAAVIFTKILN--L  56
Db  631 GSIINIAGNLIVSKGANLQAITNYTFNVAGSFDNNGASNISIARGGAK-FKDINNTSSL  688

Qy   57 PVDENAQPSFKDAKNIWSSKYIAAVEKAGVVKGDKENFYPEGKIDRASFASMLVSAYNL  116
Db  689 NITTNSDTTYR----------------TIIKGNISNKSGDLNI-----IDKKSDAEIQIGG-NI  730

Qy  117 KDKVNGELVTTFEDLLDHWGEKANLINLGISVGTGGKWEPNKSVSRAEAAQFIALTDK  176
Db  731 SQK-EGNLTIS----------SDKVNITNQIITKAGVEG-  758

Qy  177 KYGKKDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVT  236
Db  759 --GRSDSSEA---------ENANLTIQTKEL-KL-AGDLNISGFNKABITAKNGSDLTIGN  806

Qy  237 LGGKVAPNKDLTV-KVKNQSFVT----KFVVEVKKLAVEKLTFDDDRAGQAIAFKLNDE  290
Db  807 ASGGNADAKKVTFDKVKDSKISTDGHNVTLNSEVKTSNGSSNAGNDNSTGLLTISAK----  862

Qy  291 KGNADVEYLNLANHDVKFVANNLDGSPANIF-EGGEATSTGKLAVGIKQGDYKVEVQV  348
Db  863 ---DV--TVNNNVTSHKTINISAAAGNVTTKEGTTINATTGSVEVTAQNG--------  907

Qy  349 TKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLV  408
Db  908 TIKGNITSQNVTVATAENLVTTENAVINA----------TSGTVNISTKTGDIKGGIESTSGN  960

Qy  409 VGEKASLNKL-VATIAGEDKVVDPGSIKSSNHGIISVVNNYITAEAAGEATLTIKVGD  467
Db  961 VNITASGNTLKVSNITGQDVTV-------TADAGALTTAGSTISATTGNANITTKTGD  1012

Qy  468 VTKDVK--------FKVTTDSRKLVSVKANPDKLQVVQNKTLPVT--FVTTDQY  511
Db 1013 INGKVESSSGSVTLVATGATLAVGNISGNTVTITADSGKLTSTVGSTINGTNSVTTSSQS 1072

Qy  512 GDPFGANTAAIKEVLPKTG---------VVAEGGLDVVTTDSGSIGTKTIGVTGNDVGEG  562
Db 1073 GDIEGTISGNTVNTASTGDLTIGNSAKVEAKNGAATLTAESGKLTTQ----TGSSI---  1125

Qy  563 TVHFQNGNGATLGSLYNVTEGNVAFKNFELVSKVGQGSPDTKLDLNVST-TVEYQLS  621
Db 1126 ----TSSNGQTTLTAKDSSIAGNINAANVTL-NTTGTLTTTGDSKINATSGTLTINAKDA 1180

Qy  622 KY-----TSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVTGKTPGKVDIH---LTKN  674
```

```
Qy  341 DYKVEVQVTKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDF  400
Db  361 TADVTAKVLPDGVLTNFKVTVTEVPV---QVQNQGFTL-VDN-------LSNA--  405

Qy  401 ALNSQNLIVGEKASLNKLVATIAGEDKVV-------DPGS---ISIKSSNHGIIS  445
Db  406 ---PQNTVAFNKA--EKVTSMFAGETKTVAMIDTKNGDPETKPVDFKDATVRSLNPIIAT  460

Qy  446 VVNN---YITAEA-AGEATLTIKVGDVTK-------DVK------FKVTTDSRKLVSVK  487
Db  461 AAINGSELLVTANAGQSGKASFEVTPKDNTKRTFTVDVKKDPVLQDIKVDATSVKLSDEA  520

Qy  488 ANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVVTTDSGSI  547
Db  521 VGGGEVEGVNQKIIKVSAV-DQYG--------KEI-KFGT--KGKVTVTTNTEGL  563

Qy  548 GTKTIGVTGNDVGEGTVHFQNGNGATLGSLYV---NVTEGNVAFKNPELVS--------  595
Db  564 VIKNVN------SDNTIDFDSGNSATDQFVVATKDKIVNGKVEVKIFKNASDTTPTSTK  617

Qy  596 -------------KVGQYGQSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEGYEVE  641
Db  618 TITVNVVNVKADATPVGLDIVAP-SEIDVNAPNTAS------TADVDFINFESVEITVTLD  670

Qy  642 S-----KNLAVADAKIVGNK--VVVTGK----------TPGKVDIHLTKNGAT---  677
Db  671 SNGNRLKKVTPTATTLVGTNDYEVNGNVLQFKGNDELTLLTSSSTVNVDVTADGITKRI  730

Qy  678 ---AGKATVEIVQETIAIKSVNFKPVQTEN--FVEKKINIGTVLELEKSNLDDIVKGI  730
Db  731 PVKYINSASVP-ASATVATSPVTVKLNSSDNDLTFEELIFGVIDPTQLVKD--EDINEFI  787

Qy  731 NLTKETQH----------KVRVVKSGAE------QGKLYLDRN---------G  758
Db  788 AVSKAAKNDGYLYNKPLVTVKDASGKVIPTGANVYGLNHDATNGNIWFDEEQAGLAKKFS  847

Qy  759 DAVFN----AGDVKLGDVTVSQTSDSALPNFKADLYDTLTTKYTDKG----TLVFK--  806
Db  848 DVHFDVDFSLANVVKTGSGTVSSS--------PSLSDAI-QLTNSGDAVSPTLVIKSI  896

Qy  807 VLKDKDVITSEIGSQAVHVNV  827
Db  897 YVKGADKDDNNLLAAPVSVNV  917
```

RESULT 5
US-08-728-470-9
; Sequence 9, Application US/08728470
; Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:

Db   117 KDKVNGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDK 176
Qy   731 SQK-EGNLTIS---------SDKVNITNQITIKAGVEG------------------ 758

Db   177 KYGKKDNAQAYVTDKVSEPTKLTITGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVT 236
Qy   759 --GRSDSSEA-------ENANLIQTKEL-KL-AGDLNISGFNKAETAKNGSDLTIGN 806

Db   237 LGGKVAPNKDLTV-KVKNQSFVT----KPVYEVKKLAVEKLTPDDRAGQAIAFKLNDE 290
Qy   807 ASGGNADAKKVTFDKVKDSKISTDGHNVLNSEVKTSNGSSNAGNDNSTGLTISAK---- 862

Db   291 KGNADVEYLNLANHDVKFVANNLDGSPANIF--EGGEATSTTGKLAVGIKQGDYKVEVQV 348
Qy   863 ----DV---TVNNNVTSHKTINISAAAGNVTTKEGTTINATTGSVEVTAQNG------- 907

Db   349 TKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLV 408
Qy   908 TIKGNITSQNVTVTATENLVTTENAVINA------TSGTVNISTKTGDIKGGIESTSGN 960

Db   409 VGEKASLNKL-VATIAGEDKVVDPGSIKSSNFGIISVVNNYITAEAAGEATLTIKVGD 467
Qy   961 VNITASGNTLKVSNITGQDVTV------TADAGALTTTAGSTISATTGNANITTKTGD 1012

Db   468 VTKDVK-------------FKVTTDSRKLJVSVKANPDKLQVVQNKTLPVT--FVTTDQY 511
Qy 1013 INGKVESSSGSVTLVATGATLAVGNISGNTVTTTADSGKLTSTVGSTINGTNSVTTSSQS 1072

Db   512 GDPFGANTAAIKEVLPKTG---------VVAEGGLDVVTTDSGSIGTKTIGVTGNDVGEG 562
Qy 1073 GDIEGTISGNTVNVTASTGDLTIGNSAKVEAKNGAATLTAESGKLTTQ---TGSSI--- 1125

Db   563 TVHFQNGNGATLGSLYVNVTBGNVAFKNFELVSKVGQYGQSPDTKLDLNVST-TVEYQLS 621
Qy 1126 ---TSSNGQTTLTAKDSSIAGNINAANVTL-NTTGTLTTTGDSKINATSGTLITINAKDA 1180

Db   622 KY----TSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVTGKTPGKVDIH---LTKN 674
Qy 1181 KLDGAASGDRTVVNATNASG-------------SGN---VTAKTSSSVNITGDLNTIN 1222

Db   675 G---ATAGKATVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLELEKSNLDDIVK-G 729
Qy 1223 GLNIISENGRNTVRLRGKEIDVKYIQPGVASVEBVIEAKRVLEKVKDLSDEERETLAKLG 1282

Db   730 IN--------LTKETQHKVRVVKSG---AEBQGKLYLDRNGDAVFNAGD 766
Qy 1283 VSAVRFVEPNNAITVNTQNEFTTKPSSQVTISEGK------ACFSSGN 1324

RESULT 7
US-08-617-697-9
; Sequence 9, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
;   APPLICANT: Barenkamp, Stephen J
;   TITLE OF INVENTION: High Molecular Weight Surface Proteins
;   TITLE OF INVENTION: of No. 597336-Typeable Haemophilus
;   NUMBER OF SEQUENCES: 11
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Shoemaker and Mattare, Ltd.
;   STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;   STREET: Bldg. 1
;   CITY: Arlington
;   STATE: Virginia
;   COUNTRY: U.S.A.
;   ZIP: 22202-0286
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/617,697

Db 1181 KLDGAASGDRTVVNATNASG-------------SGN---VTAKTSSSVNITGDLNTIN 1222
Qy   675 G---ATAGKATVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLELEKSNLDDIVK-G 729

Db 1223 GLNIISENGRNTVRLRGKEIDVKYIQPGVASVEBVIEAKRVLEKVKDLSDEERETLAKLG 1282
Qy   730 IN--------LTKETQHKVRVVKSG---AEQGKLYLDRNGDAVFNAGD 766

Db 1283 VSAVRFVEPNNAITVNTQNEFTTKPSSQVTISEGK------ACFSSGN 1324

RESULT 6
US-08-719-641-9
; Sequence 9, Application US/0871964l
; Patent No. 6218141
; GENERAL INFORMATION:
;   APPLICANT: Barenkamp, Stephen J
;   TITLE OF INVENTION: High Molecular Weight Surface Proteins
;   TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
;   NUMBER OF SEQUENCES: 10
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Shoemaker and Mattare, Ltd.
;   STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;   STREET: Bldg. 1
;   CITY: Arlington
;   STATE: Virginia
;   COUNTRY: U.S.A.
;   ZIP: 22202-0286
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/719,641
;   FILING DATE:
;   CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/302,832
;   FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US PCT/US93/02166
;   FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: GB 9205704.1
;   FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
;   NAME: Berkstresser, Jerry W
;   REGISTRATION NUMBER: 22,651
;   REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703) 415-0810
;   TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 9:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1338 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
US-08-719-641-9

Query Match   6.1%; Score 257.5; DB 3; Length 1338;
Best Local Similarity   22.0%; Pred. No. 1.8e-10;
Matches 183; Conservative 104; Mismatches 328; Indels 215; Gaps 38;

Qy   16 GSI-----NYLVDKGA--------ITGKPDGTYGPTESIDRASAAVIFTKILN--L 56
Db   631 GSININIAGNLTVSKGANLQAITNYTFNVAGSFDNNGASNISIARGGAK-FKDINNTSSL 688

Qy   57 PVDENAQPSFKDAKNIWSSKYIAAVEKAGVVKGDKENFYPEGKIDRASPASMLVSAYNL 116
Db   689 NITTNSDTTYR---------TIIKGNISNKSGDLNI-----IDKKSDAEIQIGG-NI 730

```
;   FILING DATE:       01-APR-1996
;   CLASSIFICATION:    424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/302,832
;   FILING DATE:       05-OCT-1994
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US PCT/US93/02166
;   FILING DATE:       16-MAR-1993
;   ATTORNEY/AGENT INFORMATION:
;   NAME:  Berkstresser, Jerry W
;   REGISTRATION NUMBER: 22,651
;   REFERENCE/DOCKET NUMBER: 1038-557
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  (703) 415-0810
;   TELEFAX:    (703) 415-0813
;   SEQUENCE CHARACTERISTICS:
;   INFORMATION FOR SEQ ID NO: 9:
;   LENGTH:  1599 amino acids
;   TYPE:  amino acid
;   STRANDEDNESS:  single
;   TOPOLOGY:  linear
US-08-617-697-9

Query Match          6.1%;     Score 257.5;  DB 2;   Length 1599;
Best Local Similarity 22.0%;   Pred. No. 2.4e-10;
Matches  183;  Conservative  104;  Mismatches  328;  Indels  215;  Gaps   38;

Qy     16 GSI-----NYLVDKGA--------ITGKPDGTYGPTESIDRASAAVIFTKILN--L     56
          |||              |    ||    ::: | : :| |  |||
Db    891 GSIINIAGNLTVSKGANLQAITNYTFNVAGSFDNNGASNISIARGGAK--FKDINNTSSL  948

Qy     57 PVDENAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNL 116
           ||   :::              : :| |: |  |    ||||
Db    949 NITTNSDTTYR-----------------TIIKGNISNKSGDLNI-----IDKKSDAEIQIGG-NI  990

Qy    117 KDKVNGELVTTFEDLLDHWGEEKANILINLGISVGTGKWEPNKSVSRAERAAQFIALTDK 176
           ::      |||      |    |||
Db    991 SQK-EGNLTIS---------SDKVNITNQITIKAGVEG-----------------------------  1018

Qy    177 KYGKKDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVT 236
           : |::| |       | ||   |    ::|::|
Db   1019 --GRSDSSEA--------ENANLTIQTKEL--KL-AGDLNISGFNKAEITAKNGSDLITIGN 1066

Qy    237 LGGKVAPNKDLTV-KVKNQSFVT----KFVYEVKKLAVEKLTFDDDRAGQAIAFKLNDE 290
           |||                |||          : : |  ||:
Db   1067 ASGGNADAKKVTFPDKVKDSKISTDGHNVTLNSEVKTSNGSSNAGNDNSTGLTISAK-----  1122

Qy    291 KGNADVEYLNLANHDVKFVANNLDGSPANIF--EGGEATSTTGKLAVGIKQGDYKVEVQV 348
           | | :::  | |         ::|   |          |  ::| |    |  | |
Db   1123 ---DV---TVNNNVTSHKTINISAAAGNVTTKEGTTINATTGSVEVTAQNG---------  1167

Qy    349 TKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLV 408
           ::| | |                 :|:   :: |:
Db   1168 TIKGNITSQNVTVTATENLVTTENAVINA------TSGTTVNISTKTGDIKGGIESTSGN  1220

Qy    409 VGEKASLNKL-VATIAGEDKVDPGSISIKSSNHGIISVVNNYITAERAGEATLTIKVGD 467
           | |:| |        |  |  |  : :::    || |   ::  : ::::|
Db   1221 VNITASGNTLKVSNITGQDVTV-------TADAGALTTTAGSTISATTGNANITTKTGD 1272

Qy    468 VTKDVK----------------FKVTTDSRKLVSVKANPDKLQVVQNKTLPVT--FVTTDQY 511
                                   | | | | : |    |  |  :::|: | :
Db   1273 INGKVESSSGSVTLVATGATLAVGNISGNTVTITADSGKLTSTVGSIINGTNSVTTSSQS 1332

Qy    512 GDPFGANTAIKEVLPKTG--------VVAEGGLDVVTTDSGSIGTKTIGVTGNDVGEG 562
                      :|:|         :|: |:| || :|| | : | | : |
Db   1333 GDIEGTISGNTVNTASTGDLITIGNSAKVEAKNGAATLTAESGKLTTQ----TGSSI---- 1385

Qy    563 TVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQYGQSPDTKLDLNVST-TVEYQLS 621
           | :   ||  | ||   :    |:   || || |: |  |  || ||:::|
Db   1386 ---TSSNGQTTLTAKDSSIAGNTNAANVTL-NTTGLTTTGDSKINATSGTLTINAKDA 1440

Qy    622 KY-----TSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVDIH--LTKN 674
           ::           |     : |  : || || | :|   |  |  | |  |    | ||
Db   1443 (cut)
```

```
Db   1441 KLDGAASGDRTVVNATNASG-----------------SGN--VTAKTSSSVNITGDLNTIN 1482

Qy    675 G----ATAGKATVEIVQSTIAIKSVNFKPVQTENFVEKKINIGTVLELEKSNLDDIVK-G 729
        |             |  |:|    ::   | |          | |  ||    :|:
Db   1483 GLNIISENGRNTVRLRGKEIDVKYIQPGVASVEEVIEAKRVLEKVKDLSDEERETLAKLG 1542

Qy    730 IN---------LTKETQHKVRVVKSG--AEQGKLYLDRNGDAVFNAGD 766
           ::          :  ::: :   |   ::|
Db   1543 VSAVRFVEPNNAITVNTQNEFTTKPSSQVTISEGK--------ACFSSGN 1584
```

```
RESULT 8
US-08-728-470-10
;   Sequence 10, Application US/08728470
;   Patent No. 5928651
;   GENERAL INFORMATION:
;   APPLICANT:  Barenkamp, Stephen J
;   TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;   TITLE OF INVENTION:  of No. 5928651-Typeable Haemophilus
;   NUMBER OF SEQUENCES:  10
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE:  Shoemaker and Mattare, Ltd.
;   STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;   STREET:  Bldg. 1
;   CITY:  Arlington
;   STATE:  Virginia
;   COUNTRY:  U.S.A.
;   ZIP:  22202-0286
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE:  Floppy disk
;   COMPUTER:  IBM PC compatible
;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:  PatentIn Release #1.0,  Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/728,470
;   FILING DATE:
;   CLASSIFICATION:  424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:  US 08/302,832
;   FILING DATE:  16-MAR-1993
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:  US PCT/US93/02166
;   FILING DATE:  16-MAR-1993
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:  GB 9205704.1
;   FILING DATE:  16-MAR-1992
;   ATTORNEY/AGENT INFORMATION:
;   NAME:  Berkstresser, Jerry W
;   REGISTRATION NUMBER:  22,651
;   REFERENCE/DOCKET NUMBER:  1038-633
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:   (703) 415-0810
;   TELEFAX:    (703) 415-0813
;   INFORMATION FOR SEQ ID NO:  10:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH:  1529 amino acids
;   TYPE:  amino acid
;   STRANDEDNESS:  single
;   TOPOLOGY:  linear
US-08-728-470-10

Query Match          6.0%;     Score 251.5;  DB 2;   Length 1529;
Best Local Similarity 22.5%;   Pred. No. 6.3e-10;
Matches  188;  Conservative  121;  Mismatches  345;  Indels  181;  Gaps   38;

Qy     16 GSINYLVDKGAITGKPDGTYGPTE-------------SIDRASAAVIFTKILNLPV     58
          |||             |   |               :||||||| ||||:|||
Db    778 GNIN-ITNKANVTLQADTSNSNTGLKKRTLTLGNISVEGNLSLTGANANIVGN--LSIAE  834

Qy     59 DENAQPSFKDAKNI--WSSKYIAAVE-KAGVVKGDGKENFYPEGKIDRASFASMLVSAY 114
          : |  |  || :   | :| :||    |  |  |||  :  |:|| |
Db    835 DSTFKGEASDNLNITGFTNNGTANINIKQGVKLQGDIN--NKGGLNITTNAS-----G  887
```

Qy   115 NLKDKVNGELVTTFEDLLDHWGEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALT 174

Db   888 TQKTIINGNITNEKGDL--NIKNIKADAEIQIG---GNISQKEGNLTISSDKVNITNQIT 942

Qy   175 DK--KYGKKDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGT 231

Db   943 IKAGVEGGRSDSSEA------ENANLTIQTKEL-KL-AGDLNISGFNKAEITAKNGSD 992

Qy   232 SAVVTLGGKVAPNKDLTV-KVKNQSFVT-----KFVYEVKKLAVEKLTFDDDRAGQAIAF 285

Db   993 LTIGNASGGNADAKKVTFDKVKDSKISTDGHNVTLNSEVKTSNGSSNAGNDNSTGLTISA 1052

Qy   286 KLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIF--EGGEATSTTGKLAVGIKQGDYK 343

Db   1053 K-------DV--TVNNNVTSHKTINISAAAGNVTTKEGTTINATTGSVEVTAQNG--- 1098

Qy   344 VEVQVTKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALN 403

Db   1099 ------TIKGNITSQNVTVTATENLVTTENAVINA------TSGTVNISTKTGDIKGGIE 1146

Qy   404 SQNLVVGEKASLNKL-VATIAGEDKVVDPGSISIKSSNHGIISVVNNYITAEAAGEATLT 462

Db   1147 STSGNVNITASGNTLKVSNITGQDVTV-------TADAGALTTTAGSTISATTGNANIT 1198

Qy   463 IKVGDVTKDVK---------FKVTTDSRKLVSVKANPDKLQVVQNKTLPVT--FV 506

Db   1199 TKTGDINGKVESSSGGSVTLVATGATLAVGNISGNTVITTADSGKLTSTVGSTINGTNSVT 1258

Qy   507 TTDQYGDPFGANTAAIKEVLPKTG--------VVAEGGLDVVTTDSGSIGTKTIGVTGN 557

Db   1259 TSSQSGDIEGTISGNTVNVTASTGDLTIGNSAKVEAKNGAATLTAESGKLTTQ----TGS 1314

Qy   558 DVGEGTVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQYGQSPDTKLDLNVST-TV 616

Db   1315 SI------TSSNGQTTLTAKDSSIAGNINAANVTL-NTTGTLTTTGDSKINATSGTLTI 1366

Qy   617 EYQLSKY---TSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVDIH-- 670

Db   1367 NAKDAKLDGAASGDRTVNATNASG----------SGN---VTAKITSSSVNITGD 1408

Qy   671 -LTKNG----ATAGKATVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLELEKSNLDD 725

Db   1409 LNTINGLNIISENGRNTVRLRGKEIDVKYIQPGVASVEEVIEAKRVLEKVKDLSDEERET 1468

Qy   726 IVK-GIN------LTKETQHKVRVVKSG--AEQGKLYLDRNGDAVFNAGD 766

Db   1469 LAKLGVSAVRFVEPNNAITVNTQNEFTTKPSSQVTISEGK-------ACFSSGN 1515

FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
    FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
    FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
    FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
  NAME: Berkstresser, Jerry W
  REGISTRATION NUMBER: 22,651
  REFERENCE/DOCKET NUMBER: 1038-625
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (703) 415-0810
  TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1529 amino acids
    TYPE: amino acid
    STRANDEDNESS: single
    TOPOLOGY: linear
US-08-719-641-10

Query Match   6.0%;   Score 251.5; DB 3;   Length 1529;
Best Local Similarity 22.5%;   Pred. No. 6.3e-10;
Matches   188; Conservative 121; Mismatches 345; Indels 181; Gaps 38;

Qy   16   GSINYLVDKGAITGKPDGTYGPTE------------------SIDRASAAVIFTKILNLPV 58

Db   778 GNIN-ITNKANVTLQADTSNSNTGLKKRTLTLGNISVEGNLSLTGANANIVGN--LSIAE 834

Qy   59   DENAQPSFKDAKNI--WSSKYIAAVE-KAGVVKGDGKENFYPEGKIDRASFASMLVSAY 114

Db   835 DSTFKGEASDNLNITGTFTNNGTANINIKQGVVKLQGDIN--NKGGLNITTNAS------G 887

Qy   115 NLKDKVNGELVTTFEDLLDHWGEKANILNLGISVGTGGKWEPNKSVSRAEAAQFIALT 174

Db   888 TQKTIINGNITNEKGDL--NIKNIKADAEIQIG---GNISQKEGNLTISSDKVNITNQIT 942

Qy   175 DK--KYGKKDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGT 231

Db   943 IKAGVEGGRSDSSEA------ENANLTIQTKEL-KL-AGDLNISGFNKAEITAKNGSD 992

Qy   232 SAVVTLGGKVAPNKDLTV-KVKNQSFVT-----KFVYEVKKLAVEKLTFDDDRAGQAIAF 285

Db   993 LTIGNASGGNADAKKVTFDKVKDSKISTDGHNVTLNSEVKTSNGSSNAGNDNSTGLTISA 1052

Qy   286 KLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIF--EGGEATSTTGKLAVGIKQGDYK 343

Db   1053 K-------DV--TVNNNVTSHKTINISAAAGNVTTKEGTTINATTGSVEVTAQNG--- 1098

Qy   344 VEVQVTKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALN 403

Db   1099 ------TIKGNITSQNVTVTATENLVTTENAVINA------TSGTVNISTKTGDIKGGIE 1146

Qy   404 SQNLVVGEKASLNKL-VATIAGEDKVVDPGSISIKSSNHGIISVVNNYITAEAAGEATLT 462

Db   1147 STSGNVNITASGNTLKVSNITGQDVTV------TADAGALTTTAGSTISATTGNANIT 1198

Qy   463 IKVGDVTKDVK--------FKVTTDSRKLVSVKANPDKLQVVQNKTLPVT--FV 506

Db   1199 TKTGDINGKVESSSGGSVTLVATGATLAVGNISGNTVITTADSGKLTSTVGSTINGTNSVT 1258

Qy   507 TTDQYGDPFGANTAAIKEVLPKTG------VVAEGGLDVVTTDSGSIGTKTIGVTGN 557

Db   1259 TSSQSGDIEGTISGNTVNVTASTGDLTIGNSAKVEAKNGAATLTAESGKLTTQ----TGS 1314

Qy   558 DVGEGTVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQYGQSPDTKLDLNVST-TV 616

Db   1315 SI-----TSSNGQTTLTAKDSSIAGNINAANVTL-NTTGTLTTTGDSKINATSGTLTI 1366

RESULT 9
US-08-719-641-10
; Sequence 10, Application US/08719641
; Patent No. 6218141
GENERAL INFORMATION:
  APPLICANT: Barenkamp, Stephen J
  TITLE OF INVENTION: High Molecular Weight Surface Proteins
  TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
  NUMBER OF SEQUENCES: 10
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Shoemaker and Mattare, Ltd.
    STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
    STREET: Bldg. 1
    CITY: Arlington
    STATE: Virginia
    COUNTRY: U.S.A.
    ZIP: 22202-0286
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/719,641

```
Qy  116 L-KDKVNGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALT 174
            :  ||:    :     :     |:   ||:|:  :     ::      :
Db  958 TQKTIINGNITNEKGDL-NIKNIKADAEIQIG---GNISQKEGNLTISSDKVNITNQIT 1012

Qy  175 DK---KYGKKDNAQAYVTDVKVSEPTKLILTGTGLDKLSADDVTLEGDKAVAIEASTDGT 231
        |:         :   :          :   |  |||:|:|:|::|  |||||::|::||
Db 1013 IKAGVEGGRSDSSEA------ENANLITQTKEL-KI-AGDLNISGFNKAEITAKNGSD 1062

Qy  232 SAVVTLGGKVAPNKDLTV-KVKNQSFVT---KPVYEVKKLAVEKLTFDDDRAGQAIAF 285
        ::     |:  |       ||||| |       |||||||||||||  |||:    |:
Db 1063 LTIGNASGGNADAKKVTFDKVKDSKISTDGHNVTLNSEVKTSNGSSNAGNDNSTGLTISA 1122

Qy  286 KLNDEKGNADVEYLNLANHDVKFVANNLDGSPANIF-EGGEATSTGKLAVGIKQGDYK--- 343
        |:             : |     :|:      :     |  :| :|||:||:||||||
Db 1123 K------DV--TVNNNVTSHKTINISAAAGNVTTKEGTTINATTGSVEVTAQNG---- 1168

Qy  344 VEVQVTKRGGLTVSNTGIITVKNLDTPASAIKNVVFALDADNDGVNYGSKLSGKDFALN 403
              :                :  :       |  ||:
Db 1169 ----TIKGNITSQNVTVTATENLVTTENAVINA------TSGTVNISTKTGDIKGGIE 1216

Qy  404 SQNLVVGEKASLNKL-VATIAGEDKVVDPGSISIKSSNHGIISVVNNYITAEAAGEATLT 462
         :||:|  |          :          :         |:
Db 1217 STSGNVNITASGNTLKVSNITGQDVTV-------TADAGALTTTAGSTISATTGNANIT 1268

Qy  463 IKVGDVTKDVK-------FKVTDSRKLVSVKANPDKLQVVQNKTLPVT--FV 506
         |  |       :      |:|   |:|:
Db 1269 TKTGDINGKVESSSGSGVTLVATGATLAVGNISGNTVTITADSGKLTSTVGSTINGTNSVT 1328

Qy  507 TTDQYGDPFGANTAAIKEVLPKTG-----VVAEGGLDVVTTDSGSIGTKTIGVTGN 557
         ::|       :         :  ::       |     :  |:     : ||  ||:
Db 1329 TSSQSGDIEGTISGNTVNVTASTGDLTIGNSAKVEAKNGAATLIAESGKLTTQ----TGS 1384

Qy  558 DVGEGTVHFGNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQYGQSPDTKLDLNVST-TV 616
                |         |        :          |     :         |::|  |
Db 1385 SI------TSSNGQTLTAKDSSIAGNINAANVTL-NTTGTLTTTGDSKINATSGTLTI 1436

Qy  617 EYQLSKY----TSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVDIH- 670
         :      |:: :          ::                |
Db 1437 NAKDAKLDGAASGDRTVVNATNASG-------SGN---VTAKTSSSVNITGD 1478

Qy  671 -LTKNG---ATAGKATVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLELEKSNLDD 725
          |  ||   ||    ||:             :        ::    |  | :  |:||
Db 1479 LNTINGLNIISENGRNTVRLRGKEIDVKYIQPGVASVEEVIEAKRVLEKVKDLSDEERET 1538

Qy  726 IVK-GIN------LTKETQHKVRVVKSG--AEQGKLYLDRNGDAVFNAGD 766
         :   |:|      |:|||||:|||||:||  |||||:|:             :
Db 1539 LAKLGVSAVRFVEPNNAITVNTQNEFTTKPSSQVTISEGK------ACFSSGN 1585
```

RESULT 11
US-09-377-155-33
; Sequence 33, Application US/09377155
; Patent No.6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/377,155
; CURRENT FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-377-155-33

```
Qy  617 EYQLSKY----TSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVTGKTPGKVDIH-- 670
         :      |:: :          ::                |
Db 1367 NAKDAKLDGAASGDRTVVNATNASG------SGN---VTAKTSSSVNITGD 1408

Qy  671 -LTKNG---ATAGKATVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLELEKSNLDD 725
          |  ||   ||    ||:             :        ::    |  | :  |:||
Db 1409 LNTINGLNIISENGRNTVRLRGKEIDVKYIQPGVASVEEVIEAKRVLEKVKDLSDEERET 1468

Qy  726 IVK-GIN------LTKETQHKVRVVKSG--AEQGKLYLDRNGDAVFNAGD 766
         :   |:|      |:|||||:|||||:||  |||||:|:             :
Db 1469 LAKLGVSAVRFVEPNNAITVNTQNEFTTKPSSQVTISEGK------ACFSSGN 1515
```

RESULT 10
US-08-617-697-10
; Sequence 10, Application US/08617697
; Patent No.5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-617-697-10

```
Query Match              5.9%;  Score 249.5;  DB 2;  Length 1600;
Best Local Similarity   22.2%;  Pred. No. 9.6e-10;
Matches 185;  Conservative 121;  Mismatches 346;  Indels 183;  Gaps 37;

Qy   16 GSINYLVDKGAITGKPDGTYGPTE-------------SIDRASAAVIFTKILNLPV 58
         :|| : |  |   :    |  ||              |: |:|||::    |||:
Db  850 GNIN-ITNKANVLQADTSNSNTGLKKRTLTLGNISVEGNLSLIGNANIVGN--LSIAE 906

Qy   59 DENAQPSFKDAKNI--WSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYN 115
         :         |::   ||| :||| :|||||                   :|  ||::
Db  907 DSTFKGEASDNLNITGTFNNGTANINIKGVVK-------LGDINNKGGLNITTNASG 957
```

GENERAL INFORMATION:
```
     APPLICANT: St. Geme, Joseph
     APPLICANT: Barenkamp, Stephen J.
     TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
     NUMBER OF SEQUENCES: 19
     CORRESPONDENCE ADDRESS:
          ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
          STREET: Four Embarcadero Center, Suite 3400
          CITY: San Francisco
          STATE: California
          COUNTRY: United States
          ZIP: 94111-4187
     COMPUTER READABLE FORM:
          MEDIUM TYPE: Floppy disk
          COMPUTER: IBM PC compatible
          OPERATING SYSTEM: PC-DOS/MS-DOS
          SOFTWARE: PatentIn Release #1.0, Version #1.30
     CURRENT APPLICATION DATA:
          APPLICATION NUMBER: US/08/913,942
          FILING DATE: 29-DEC-1997
          CLASSIFICATION: 514
     PRIOR APPLICATION DATA:
          APPLICATION NUMBER: US 08/409,995
          FILING DATE: 24-MAR-1995
     PRIOR APPLICATION DATA:
          APPLICATION NUMBER: PCT/US96/4031
          FILING DATE: 22-MAR-1996
     ATTORNEY/AGENT INFORMATION:
          NAME: Vance, Dolly A.
          REGISTRATION NUMBER: 39,054
          REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV
     TELECOMMUNICATION INFORMATION:
          TELEPHONE: (415) 398-3249
          TELEFAX: (415) 781-1989
          TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
     SEQUENCE CHARACTERISTICS:
          LENGTH: 2353 amino acids
          TYPE: amino acid
          TOPOLOGY: linear
     MOLECULE TYPE: protein
US-08-913-942-4
```

```
Query Match          5.9%;  Score 249.5;  DB 3;  Length 2353;
Best Local Similarity 22.0%;  Pred. No. 1.7e-09;
Matches 201;  Conservative 107;  Mismatches 335;  Indels 269;  Gaps 45;

Qy    23 DKGAITGKPDGTYGPTESIDRASAAV--IFTKILNLPVDENAQPSFKDAKNIWSSKYIAA   80
            :|:|| | |||:
Db   705 DRGKVTVK-DAT---ANDADKKVATVKDVATAI-------NSAATFVKTENLTTS-----  748

Qy    81 VEKAGVVKGDGKENFYPEGKIDRASFASMLV--SAYNLKDKVNGELVTTFEDLLDHWGEE  138
              :|:|
Db   749 -------IDEDNPTDNGKDDALKAGDTLTFKAGKNLKVKRDGKNITF--DLAKNLEVK   797

Qy   139 KANI--LINLGISVGTGGKWEPNK--SVSRAEAAQFIALTDKKYGKKDNAQAYVTDVKVS  194
            :|: |
Db   798 TAKVSDTLTIGGNTPTGGTTATPKVNITSTADGLNFAKETADASGSKN---VYLKGI---  851

Qy   195 EPTKLTLTGTGLDKLSADDVTLEGDK---AVAIEASTDGTSAVVTLGGKVAPNKDLTVKV  251
            :|: |
Db   852 -ATTLTEPSAGA-KSSHVDLNVDATKKSNAASIE------DVLRAGWNIQGNGN-----  897

Qy   252 KNQSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLNLANHDVKFVAN  311
            :|:
Db   898 -NVDYVATY----------DTVNFTDDSTGTTTV-------------------TVTQ   924

Qy   312 NLDGSPANIFEGGEAT--STTGKLAVGIKQGDYKVEVQVTKRGGLTVS------NTGI  361
            :| :|
Db   925 KADGKGADVKIGAKTSVIKDHNGKLFTGKDLKD--------ANNGATVSEDDGKDTGTGL  976

Qy   362 ITVKNL--------------------------DTPASAIKNVVFALDADNDGVVNYGKLSGKD----399
            :| |:| |:|:|:
Db   977 VTAKTVIDAVNKSGWRVTGEGATAETGATAV------NAGNAETVTSGTSVNFKNGNAT 1029

Qy   400 -FALNSQNLVVGEKASLN----------KLVA------TIAGEDKVVDPGSISIKSSN--  440
            :|:|  |             |         ::
Db  1030 TATVSKDNGNINVKYDVNVGDGLKIGDDKKIVADTTLTVTVTGGKVSVPAGANSVNNNKKL 1089

Qy   441 ---HGIISVVNN---------YITAEAAGEATLTIKVGDVTVKDVKFKVTTDSRKLVSVK-  487
              |||                |:| ||
Db  1090 VNAEGLATALNNLSWTAKADKYADGESEGETDQEVKAGD------KVTFKAGKNLKVKQ  1142

Qy   488 ANPDKLQVVQNKTLPVTFVTFTTDQYGDPPGANTAAIKEVLPKTG-VVAEGGLDVVTTDSGS  546
              :|:| |           ||
Db  1143 SEKDFTVSLQDTLTGLTSITLG--GTANGRN----------DTGTVINKDGLTITLANGAA 1191

Qy   547 IGT-----KTIGVTGNDVGEGTVHFQNGNGATLGSLYVNV-----TEGNVAFKNFB  592
            :|:|                |             ||
Db  1192 AGTDASNGNTISVTKDGISAGNKEITNVKSAL--KTYKDTQNTADETQDKEFHAAVKNAN 1249

Qy   593 LVSKVGYGQSPDTKLDLNVSTTVEYQ----LSKYTSDRVYSDPENLEGYEVESK  643
            ||||               |||
Db  1250 EVEFVGKNGATVSAKTDNNGKHTVTIDVAEAKVGDGLEKDTDGKIKLKVDNTDG----N 1304

Qy   644 NLAVADAKIVGNKVVTGKTPGKVDIHLTKNGATAGKATVEIVQETIAIKSVN--  696
              :|: |
Db  1305 NLLTVDA------TKGASVAKGEFNAVTTPATTAQGTNANERGKVVVKGSNGATATET 1356

Qy   697 -FKPVQTENFVEKKIN-IGTVLELE---KSNLDD--------IVKGINLTKETQHKVR  741
            :|:|                             ::
Db  1357 DKKKVATVGDVAKAINDAATFVKVENDDSATIDDSPTDDGANDALKAGDTLTLKAGKNLK 1416

Qy   742 V-------VKSGAEQGKLYLDRNGDAVFNAGDVK-LGDVTVSQTSDSA--  781
            |                                  ||
Db  1417 VKRDGKNITFALANDLSVKSATVSDKLSLGTNGNKVNITSDTKGLNFAKDSKTGDDANIH 1476

Qy   782 LPNFKADLYDTL  793
            |||||
Db  1477 LNGIASTLTDTL 1488

RESULT 12
US-08-913-942-4
; Sequence 4, Application US/08913942
; Patent No. 6200578
```

```
Qy   81  VEKAGVVKGDGKENFYPEGKIDRASFASMLV--SAYNLKDKVNGELVTFEDLLDHWGEE  138
Db  749  --------IDEDNPTDNGKDDALKAGDTLTFKAGKNLKVKRDGKNITF--DLAKNLEVK  797
Qy  139  KANI--LINLGISVGTGKWEPNK--SVSRAEAAQPIALTDKKYGKDNAQAYVTDVKVS   194
Db  798  TAKVSDTLTIGGNTPTGGTTATPKVNITSTADGLNFAKETADASGSKN--VYLKGI--   851
Qy  195  EPTKLTLGTGLDKLSADDVTLEGDK--AVAIEASTDGTSAVVTLGGKVAPNKDLTVKV   251
Db  852  -ATTLTEPSAGA-KSSHVDLNVDATKKSNAASIE------DVLRAGWNIQGNGN----  897
Qy  252  KNQSFVTKFVYEVKKLAVEKLTPDDRAGQAIAFKLNDEKGNADVEYLNLANHDVKFVAN  311
Db  898  -NVDYVATY------------DTVNFTDDSTGTTTV------------------TVTQ   924
Qy  312  NLDGSPANIFEGGEAT--STTGKLAVGIKQQDYKVEVQVTKRGGLTVS------NTGI  361
Db  925  KADGKGADVKIGAKTSVIKDHNGKLFTGKDLKD------------ANNGATVSEDDGKDTGTGL  976
Qy  362  ITVKNL-----------------DTPASAIKNVVFALDADNDGVVNYGSKLSGKD---  399
Db  977  VTAKTVIDAVNKSGWRVTGEGATAETGATAV------NAGNAETVTSGTSVNFKNGNAT 1029
Qy  400  -FALNSQNLVVGEKASLN-----------KLVA-----TIAGEDKVVDPGSISIKSSN-  440
Db 1030  TATVSKDNGNINVKYDVNVGDGLKIGDDKKIVADTTTLTVTGGKVSVPAGANSVNNNKKL 1089
Qy  441  ---HGIISVVNN-------YITAEAAGEATLTIKVGDVTKDVKFKVTTDSRKLVSVK-----  487
Db 1090  VNAEGLATALNNLSWTAKADKYADGESEGETDQEVKAGD------KVTFKAGKNLKVKQ 1142
Qy  488  ANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTG-VVAEGGLDVVTTDSGS  546
Db 1143  SEKDFTYSLQDTLTGLTSITLG--GTANGRN--------DTGTVINKDGLTITLANGAA 1191
Qy  547  IGT-----KTIGVTGNDVGEGTVHFQNGNGATLGSLYVNV-------TEGNVAFKNFE   592
Db 1192  AGTDASNGNTISVTKDGISAGNKEITNVKSAL--KTYKDTQNTADETQDKEFHAAVKNAN 1249
Qy  593  LVSKVGQYGQSPDTKLDLNVSTTVEYQ---------LSKYTSDRVYSDPENLEGYEVESK  643
Db 1250  EVEFVGKNGATVSAKTDNNGKHTVTIDVAEAKVGDGLEKDTDGKIKLKVDNTDG-----N 1304
Qy  644  NLAVADAKIVGNKVVVTGKTPGKVDIHLTKNGATAGKATVEIVQETIAIKSVN-------  696
Db 1305  NLLTVDA-------TKGASVAKGEFNAVTTDATTAQGTNANERGKVVVKGSNGATATET 1356
Qy  697  -FKPVQTENFVEKKIN-IGTVLELE---KSNLDD------IVKGINLTKBTQHKVR    741
Db 1357  DKKKVATVGDVAKAINDAATFVKVENDDSATIDDSPTDDGANDALKAGDTLTLKAGKNLK 1416
Qy  742  V------------VKSGAEQGKLYLDRNGDAVFNAGDVK-LGDVTVSQTSDSA---    781
Db 1417  VKRDGKNITFALANDLSVKSATVSDKLSLGTNGNKVNITSDTKGLNFAKDSKTGDDANIH 1476
Qy  782  LPNFKADLYDTL  793
Db 1477  LNGIASTLTDTL 1488
```

RESULT 14
US-09-797-862-33
; Sequence 33, Application US/09797862
; Patent No. 6607729
; GENERAL INFORMATION:
; APPLICANT: PEAK, IAN RICHARD ANSELM
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: MOXON, E. RICHARD
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0134
; CURRENT APPLICATION NUMBER: US/09/797,862
; CURRENT FILING DATE: 2001-05-03

```
Db  977  VTAKTVIDAVNKSGWRVTGEGATAETGATAV------NAGNAETVTSGTSVNFKNGNAT 1029
Qy  400  -FALNSQNLVVGEKASLN-----------KLVA-----TIAGEDKVVDPGSISIKSSN-  440
Db 1030  TATVSKDNGNINVKYDVNVGDGLKIGDDKKIVADTTTLTVTGGKVSVPAGANSVNNNKKL 1089
Qy  441  ---HGIISVVNN-------YITAEAAGEATLTIKVGDVTKDVKFKVTTDSRKLVSVK-----  487
Db 1090  VNAEGLATALNNLSWTAKADKYADGESEGETDQEVKAGD------KVTFKAGKNLKVKQ 1142
Qy  488  ANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTG-VVAEGGLDVVTTDSGS  546
Db 1143  SEKDFTYSLQDTLTGLTSITLG--GTANGRN--------DTGTVINKDGLTITLANGAA 1191
Qy  547  IGT-----KTIGVTGNDVGEGTVHFQNGNGATLGSLYVNV-------TEGNVAFKNFE   592
Db 1192  AGTDASNGNTISVTKDGISAGNKEITNVKSAL--KTYKDTQNTADETQDKEFHAAVKNAN 1249
Qy  593  LVSKVGQYGQSPDTKLDLNVSTTVEYQ---------LSKYTSDRVYSDPENLEGYEVESK  643
Db 1250  EVEFVGKNGATVSAKTDNNGKHTVTIDVAEAKVGDGLEKDTDGKIKLKAGKNLK      1304
Qy  644  NLAVADAKIVGNKVVVTGKTPGKVDIHLTKNGATAGKATVEIVQETIAIKSVN-------  696
Db 1305  NLLTVDA-------TKGASVAKGEFNAVTTDATTAQGTNANERGKVVVKGSNGATATET 1356
Qy  697  -FKPVQTENFVEKKIN-IGTVLELE---KSNLDD                          741
Db 1357  DKKKVATVGDVAKAINDAATFVKVENDDSATIDDSPTDDGANDALKAGDTLTLKAGKNLK 1416
Qy  742  V------------VKSGAEQGKLYLDRNGDAVFNAGDVK-LGDVTVSQTSDSA---    781
Db 1417  VKRDGKNITFALANDLSVKSATVSDKLSLGTNGNKVNITSDTKGLNFAKDSKTGDDANIH 1476
Qy  782  LPNFKADLYDTL  793
Db 1477  LNGIASTLTDTL 1488
```

RESULT 13
US-09-669-974-33
; Sequence 33, Application US/09669974
; Patent No. 6333173
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/669,974
; CURRENT FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: US 09/377,155
; PRIOR FILING DATE: 1999-08-19
; PRIOR APPLICATION.NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-669-974-33

Query Match   5.9%;   Score 249.5; DB 4; Length 2353;
Best Local Similarity 22.0%;   Pred. No. 1.7e-09;
Matches 201; Conservative 107; Mismatches 335; Indels 269; Gaps 45;

```
Qy   23  DKGAITGKPDGTYGPTESIDRASAAV--IFTKILNLPVDENAQPSFKDAKNIWSSKYIAA   80
Db  705  DRGKVTVK-DAT--ANDADKKVATVKDVATAI------NSAATFVKTENLTTS------  748
```

```
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-797-862-33

Query Match        5.9%;      Score 249.5;  DB 4;   Length 2353;
Best Local Similarity  22.0%;  Pred. No. 1.7e-09;
Matches  201;  Conservative 107;  Mismatches 335;  Indels 269;  Gaps  45;

Qy    23  DKGAITGKPDGTYGPTESIDRASAAV--IFTKILNLPVDENAQPSFKDAKNIWSSKYIAA   80
Db   705  DRGKVTVK-DAT---ANADKKVATVKDVATAI------NSAATFVKTENLTTS-----   748
Qy    81  VEKAGVVKGDGKENFYPEGKIDRASFASMLV--SAYNLKDKVNGELVTTPEDLLDHWGEE  138
Db   749  ------IDEDNPTDNGKDDALKAGDTLTFKAGKNLKVRDGKNITF--DLAKNLEVK     797
Qy   139  KANI--LINLGISVGTGGKWEPNK--SVSRAEAAQFIALTDKKYGKKDNAQAYVTDVKVS  194
Db   798  TAKVSDTLTIGGNTPFGGTTATPKVNITSTADGLNFAKETADASGSKN---VYLKGI--   851
Qy   195  EPTKLTLTGTGLDKLSADDVTLEGDK---AVAIEASTDGTSAVVTLGGKVAPNKDLTVKV  251
Db   852  -ATTLTEPSAGA-KSSHVDLNVDATKKSNAASIE-------DVLRAGWNIQGNGN----   897
Qy   252  KNQSFVTKFPVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLNLANHDVKFVAN  311
Db   898  -NVDYVATY--------------DTVNFTDDSTGTTTV-----------------------  924
Qy   312  NLDGSPANIFEGGEAT--STTGKLAVGIKQGDYKVEVQVTKRGGLTVS-------NTGI   361
Db   925  KADGKGADVKIGAKTSVIKDHNGKLFTGKDLKD-------ANNGATVSEDDGKDTGTGL   976
Qy   362  ITVKNL-------------------DTPASAIKNVVFALDADNDGVVNYGSKLSGKD---  399
Db   977  VTAKTVDAVNKSGWRVTGEGATAETGATAV----------NAGNAETVTSGTSVNFKNGNAT 1029
Qy   400  -FALNSQNLVVGEKASLN-----------KLVA------TIAGEDKVVDPGSISIKSSN--  440
Db  1030  TATVSKDNGNINVKYDVNVGDGLKIGDDKKIVADTTLTVTGGKVSVPAGANSVNNNKKL  1089
Qy   441  ---HGIISVVNN----------YITAEAAGEATLTIKVGDVTKDVKFKVTTDSRKLVSVK-  487
Db  1090  VNAEGLATALNNLSWTAKADKYADGESEGETDQEVKAGD--------KVTFKAGKNLKVKQ 1142
Qy   488  ANPDKLQVVQNKTLPVTFVTVTDQYGDPFGANTAAIKEVLPKTG-VVAEGGLDVVTTDSGS  546
Db  1143  SEKDFTYSLQDTLTGLTSITLG--GTANGRN------------DTGTVINKDGLTITLANGAA 1191
Qy   547  IGT-----KTIGVTGNDVGEGTVHFQNGNGATLGSLYVNV-------TEGNVAFKNFE   592
Db  1192  AGTDASNGNTISVTKDGISAGNKEITNVKSAL--KTYKDTQNTADETQDKEFHAAVKNAN  1249
Qy   593  LVSKVGQYGQSPDTKLDLNVSTTVEYQ-------LSKYTSDRVYSDPENLEGYEVESK   643
Db  1250  EVEBFVGKNGATVSAKTDNNGKHTVTIIDVARAKVGDGLEKDTDGKIKLKVDNTDG----N 1304
Qy   644  NLAVADAKIVGNKVVVTGKTPGKVDIHLTKNGATAGKATVEIVQETIAIKSVN------  696
Db  1305  NLLTVDA--------TKGASVAKGEFNAVTTDATTAQGTNANERGKVVVKGSNGATATET 1356
Qy   697  -FKPVQTENFVEKKIN-IGTVLELE---KSNLDD------IVKGINLTKETQHKVR   741
Db  1357  DKKKVATVGDVAKAINDAATFVKVENDDSATTIDDSPTDDGANDALKAGDTLITLKAGKNLK 1416
Qy   742  V-----------VKSGAEQGKLYLDRNGDAVPNAGDVK-LGDVTVSQTSDSA--   781
Db  1417  VKRDGKNITPALANDLSVKSATVSDKLSLSGTNGNKVNITSDTKGLNFAKDSKTGDDANIH 1476
Qy   782  LPNFKADLYDTL   793
Db  1477  LNGIASTLIDTL  1488
```

RESULT 15
US-09-268-347-47
```
; Sequence 47, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 2354
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-47

Query Match        5.9%;      Score 247.5;  DB 4;   Length 2354;
Best Local Similarity  22.0%;  Pred. No. 2.5e-09;
Matches  201;  Conservative 107;  Mismatches 335;  Indels 269;  Gaps  45;

Qy    23  DKGAITGKPDGTYGPTESIDRASAAV--IFTKILNLPVDENAQPSFKDAKNIWSSKYIAA   80
Db   705  DRGKVTVK-DAT---ANADKKVATVKDVATAI------NSAATFVKTENLTTS-----   748
Qy    81  VEKAGVVKGDGKENFYPEGKIDRASFASMLV--SAYNLKDKVNGELVTTPEDLLDHWGBE  138
Db   749  ------IDEDNPTDNGKDDALKAGDTLTFKAGKNLKVRDGKNITF--DLAKNLEVK     797
Qy   139  KANI--LINLGISVGTGGKWEPNK--SVSRAEAAQFIALTDKKYGKKDNAQAYVTDVKVS  194
Db   798  TAKVSDTLTIGGNTPFGGTTATPKVNITSTADGLNFAKETADASGSKN---VYLKGI--   851
Qy   195  EPTKLTLTGTGLDKLSADDVTLEGDK---AVAIEASTDGTSAVVTLGGKVAPNKDLTVKV  251
Db   852  -ATTLTEPSAGA-KSSHVDLNVDATKKSNAASIE-------DVLRAGWNIQGNGN----   897
Qy   252  KNQSFVTKFPVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLNLANHDVKFVAN  311
Db   898  -NVDYVATY--------------DTVNFTDDSTGTTTV-----------------------  924
Qy   312  NLDGSPANIFEGGEAT--STTGKLAVGIKQGDYKVEVQVTKRGGLTVS-------NTGI   361
Db   925  KADGKGADVKIGAKTSVIKDHNGKLFTGKDLKD-------ANNGATVSEDDGKDTGTGL   976
Qy   362  ITVKNL-------------------DTPASAIKNVVFALDADNDGVVNYGSKLSGKD---  399
Db   977  VTAKTVTDAVNKSGWRVTGEGATAETGATAV----------NAGNAETVTSGTSVNFKNGNAT 1029
Qy   400  -FALNSQNLVVGEKASLN-----------KLVA------TIAGEDKVVDPGSISIKSSN--  440
Db  1030  TATVSKDNGNINVKYDVNVGDGLKIGDDKKIVADTTLTVTGGKVSVPAGANSVNNNKKL  1089
Qy   441  ---HGIISVVNN----------YITAEAAGEATLTIKVGDVTKDVKFKVTTDSRKLVSVK-  487
Db  1090  VNAEGLATALNNLSWTAKADKYADGESEGETDQEVKAGD--------KVTFKAGKNLKVKQ 1142
Qy   488  ANPDKLQVVQNKTLPVTFVTVTDQYGDPFGANTAAIKEVLPKTG-VVAEGGLDVVTTDSGS  546
Db  1143  SEKDFTYSLQDTLTGLTSITLG--GTANGRN------------DTGTVINKDGLTITLANGAA 1191
Qy   547  IGT-----KTIGVTGNDVGEGTVHFQNGNGATLGSLYVNV-------TEGNVAFKNFE   592
Db  1192  AGTDASNGNTISVTKDGISAGNKEITNVKSAL--KTYKDTQNTADETQDKEFHAAVKNAN  1249
```

```
Qy     593  LVSKVGQYGQSPDTKLDLNVSTTVEYQ--------LSKYTSDRVYSDPENLEGYEVESK  643
                 |||: .|:    |   ||        |  ||.|    :: ||    ||  :| :|
Db    1250  EVEFVGKNGATVSAKTDNNGKHTVTIDVAEAKVGDGLEKDTDGKIKLKVDNTDG------N 1304

Qy     644  NLAVADAKIVGNKVVVTGKTPGKVDIHLTKNGATAGKATVEIVQETIAIKSVN-------  696
                 ||    ||       || : : :    :   :  : ||    :|  ::|| |
Db    1305  NLLTVDA--------TKGASVAKGEFNAVTTDATTAQGTNANERGKVVKGSNGATATET  1356

Qy     697  -FKPVQTENFVEKKIN-IGTVLELE---KSNLDD----------IVKGINLTKETQHKVR  741
                  |||        |||     |    :||      : :|          ||  :  :
Db    1357  DKKKVATVGDVAKAINDAATFVKVENDDSATIDDSPTDDGANDALKAGDTLTLKAGKNLK 1416

Qy     742  V---------------VKSGAEQGKLYLDRNGDAVFNAGDVK-LGDVTVSQTSDSA----  781
                |              |||         ||    ||  |   ||    |  : |
Db    1417  VKRDGKNITFALANDLSVKSATVSDKLSLGTNGNKVNITSDTKGLKFAKDSKTGDDANIH 1476

Qy     782  LPNFKADLYDTL  793
                 |||   | |||
Db    1477  LNGIASTLTDTL 1488


Search completed: April  7, 2004, 17:34:06
Job time : 28 secs
```

```
                    GenCore version 5.1.6
         Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    April 7, 2004, 17:29:55 ; Search time 18 Seconds
                              (without alignments)
                              2409.691 Million cell updates/sec

Title:    US-09-844-281-1
Perfect score:   4202
Sequence:   1 AGKSFPDVPAGHWAEGSINY.........ITSEIGSQAVHVNVLNNPNL 833

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:   141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:      141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES

                 %
Result          Query
No.   Score   Match Length DB  ID            Description
----  -----   ----- ------ --  --            -----------
 1    4202    100.0   862   1   SLA2_BACAA    P94217 bacillus an
 2    2833.5   67.4   874   1   SLAP_BACLI    P49052 bacillus li
 3     719     17.1   814   1   SLA1_BACAA    P49051 bacillus an
 4     531     12.6   531   1   YQ42_BACAN    Q9mzo0 bacillus an
 5     348      8.3  1176   1   SLAP_BACSH    P38537 bacillus sp
 6     265.5    6.3   762   1   SLAP_ACEKI    P22258 acetogenium
 7     252.5    6.0   941   1   GUN_BACS6     P19424 bacillus sp
 8     230      5.5  1325   1   YDEK_ECOLI    P32051 escherichia
 9     223.5    5.3  2003   1   YDBA_ECOLI    P33666 escherichia
10     219      5.2  1116   1   SLPH_BRECH    P38538 brevibacill
11     217      5.2  1655   1   OMPB_RICCN    Q9kka3 r outer mem
12     211      5.0  1053   1   SLPM_BACBR    P06546 bacillus br
13     207.5    4.9  1645   1   OMPB_RICTY    P95989 r outer mem
14     207.5    4.9  2358   1   YBEJ_ECOLI    P76347 escherichia
15     207.5    4.9  2660   1   YBEJ_ECO57    Q8x8v7 escherichia
16     206.5    4.9  1654   1   OMPB_RICRI    Q53047 r outer mem
17     206      4.8  1698   1   4I_DROME      Q9v8r9 drosophila
18     200      4.8  1643   1   OMPB_RICPR    Q53020 r outer mem
19     199.5    4.7   939   1   SLAP_CAMFE    P25827 campylobact
20     197.5    4.7  1608   1   HLYA_SERMA    Q53320 serratia ma
21     195.5    4.7  1848   1   CBPA_CLOCL    P38058 clostridium
22     194.5    4.6  1656   1   OMPB_RICJA    O06653 r outer mem
23     194.5    4.6  1902   1   P2P_LACPA     Q02470 lactobacill
24     194      4.6  1300   1   120K_RICRI    P14914 rickettsia
25     193      4.6   917   1   SLAP_THETH    P35830 thermus the
26     189.5    4.5  1183   1   CNA_STAAU     Q53654 staphylococ
27     189      4.5  1953   1   BIGA_SALTY    P25927 salmonella
28     188.5    4.5  1276   1   PMP6_CHLPN    Q92899 chlamydia p
29     187.5    4.5  1861   1   APU_THETU     P38536 t amylopull
30     185.5    4.4  1577   1   HLYA_PROMI    P16466 proteus mir
31     183.5    4.4  1087   1   XYNX_CLOTM    P38535 clostridium
32     183      4.4  2249   1   OMPA_RICRI    P15921 rickettsia
33     183      4.4  4705   1   FAT2_DROME    Q9vw71 drosophila
```

```
34   181.5   4.3   1228   1   SLAP_BACST    P35825 bacillus st
35   181.5   4.3   1902   1   P2P_LACLC     P15293 lactococcus
36   180     4.3   1286   1   AIDA_ECOLI    Q03155 escherichia
37   180     4.3   1569   1   YPJA_ECOLI    P52143 escherichia
38   179.5   4.3   1726   1   MSP1_PLAFC    P04934 plasmodium
39   179.5   4.3   1726   1   MSP1_PLAFP    P50495 plasmodium
40   179     4.3   1664   1   SLP1_CLOTM    Q06852 clostridium
41   178.5   4.2   2021   1   OMPA_RICCN    Q52657 rickettsia
42   176.5   4.2    447   1   ANCA_CLOTM    Q06848 clostridium
43   176.5   4.2   1902   1   P1P_LACLC     P16271 lactococcus
44   174     4.1   1025   1   SLAP_CAUCR    P35828 caulobacter
45   172     4.1   1020   1   BCA_STRAG     Q02192 streptococc
```

                              ALIGNMENTS

```
RESULT 1
SLA2_BACAA
ID   SLA2_BACAA      STANDARD;      PRT;     862 AA.
AC   P94217;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   S-layer protein EA1 precursor.
GN   EAG OR BA0887.
OS   Bacillus anthracis (strain Ames), and
OS   Bacillus anthracis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=198094, 1392;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ames;
RX   MEDLINE=22608414; PubMed=12721629;
RA   Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA   Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA   Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA   Kolonay J.F., Beanan M.J., Dodson R.J., Helgason E., Rilstone J., Wu M.,
RA   DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA   Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA   Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA   Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA   Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA   Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA   Fraser C.M.;
RT   "The genome sequence of Bacillus anthracis Ames and comparison to
RT   closely related bacteria.";
RL   Nature 423:81-86(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sterne / 9131;
RX   MEDLINE=97260111; PubMed=9106206;
RA   Mesnage S., Tosi-Couture E., Mock M., Gounon P., Fouet A.;
RT   "Molecular characterization of the Bacillus anthracis main S-layer
RT   component: evidence that it is the major cell-associated antigen.";
RL   Mol. Microbiol. 23:1147-1155(1997).
CC   -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
CC       of proteins which coat the surface of bacteria.
CC   -!- SUBCELLULAR LOCATION: Cell wall.
CC   -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboratio
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on it
CC   use by non-profit institutions as long as its content is in no wa
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AB017027; AAP24884.1; -.
DR   EMBL; X99724; CAA68063.1; -.
DR   TIGR; BA0887; -.
DR   InterPro; IPR001119; SLH.
```

```
AC   P49052;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   S-layer protein precursor (Surface layer protein).
OS   Bacillus licheniformis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1402;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NM 105;
RX   MEDLINE=97082965; PubMed=8964497;
RA   Zhu X., McVeigh R.R., Malathi P., Ghosh B.K.;
RT   "The complete nucleotide sequence of the Bacillus licheniformis NM105
RT   S-layer-encoding gene.";
RL   Gene 173:189-194(1996).
CC   -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
CC       of proteins which coat the surface of bacteria.
CC   -!- SUBCELLULAR LOCATION: Cell wall.
CC   -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the  Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use by  non-profit  institutions  as long  as its content  is  in  no  way
CC   modified and this statement is not removed. Usage  by  and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U38842; AAC44405.1; -.
DR   PIR; JC4930; JC4930.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF00395; SLH; 3.
KW   Signal; Cell wall; S-layer; Repeat.
FT   SIGNAL        1       30    POTENTIAL.
FT   CHAIN        31      874    S-LAYER PROTEIN.
FT   DOMAIN       33       93    SLH 1.
FT   DOMAIN       94      155    SLH 2.
FT   DOMAIN      156      217    SLH 3.
SQ   SEQUENCE   874 AA;  92734 MW;  EFADCA4FF27D32AF CRC64;

Query Match             67.4%;   Score 2833.5;  DB 1;  Length 874;
Best Local Similarity   67.3%;   Pred. No. 3.6e-116;
Matches  569;  Conservative  98;  Mismatches 156;  Indels  23;  Gaps  7;

Qy    1 AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE 60
          |||||||||||||| ||||||||||| ||||||||||||||||||||||||||||||||
Db   30 AGKSFPDVPAGHWAEDSINYLVDKGAIVGKPDGTYGPTESIDRASAAVIFTKILNLPVDE 89

Qy   61 NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV 120
          ||||||||||| ||||||||||||||||||||: |||||||||||||||||: ||||:|||
Db   90 NAQPSFKDAKNLWSSKYIAAVEKAGVVKGDGKNFYPEGKIDRASFASMLVGAYNLKEKV 149

Qy  121 NGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYGK 180
          :| |||:| ||::||||||||| |||||||||| |||||||||||||||||||||||:|
Db  150 DGTLVTKFPDLRGHWGEEKANILVNLGISVGTGKWEPNKSVSRAEAAQFIALTDKKYAK 209

Qy  181 KDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVTLGGK 240
          ::. ::. . . :: ||||||||| ||||: :||||||:||:|:|||.||.||||||.||
Db  210 PENSDAKVTNVAATEPTQLTLTGTGLNKLTAEDVTLEGNKAIALEASKDGKSAVVTLSGK 269

Qy  241 VAPNKDLTVKVKNQSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLN 300
          :|||::| |||||.||:|||||||||||:||:||||||||:|:.||||||||||||:|||:
Db  270 IAPNKELPVKVKGNTFIVKVYEVKKLRVEQLTFDDDRADQAVVFKLNDEKGNADIEYLD 329

Qy  301 LANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLTVSNTG 360
          :| |||||||||||| ||||||||.| |||||||||||:|.||||||||||||||||||||
Db  330 IAGHDVKFVANNLDGTPANIFEGGTAESTTGKLAVGIAGKYKVEVQVTKRGGITVSNTG 389

Qy  361 IITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLNKLVA 420
          ||.||||| . :|||:||||:| |:|.|  |||| ||.|||||:| || ||||:|:|||
Db  390 IIEVKNLDAEATAIKDVVFAVDTDKAG-VNYAKPLSGTDFTLNSKTLVAGEKAGIHKVVA 448
```

```
DR   Pfam; PF00395; SLH; 3.
KW   Signal; Repeat; Cell wall; S-layer; Complete proteome.
FT   SIGNAL        1       29    POTENTIAL.
FT   CHAIN        30      862    S-LAYER PROTEIN EA1.
FT   DOMAIN       34       76    SLH 1.
FT   DOMAIN       95      136    SLH 2.
FT   DOMAIN      157      197    SLH 3.
SQ   SEQUENCE   862 AA;  91362 MW;  CB16B202F62CCCA0 CRC64;

Query Match             100.0%;   Score 4202;  DB 1;  Length 862;
Best Local Similarity   100.0%;   Pred. No. 1.3e-175;
Matches  833;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   30 AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE 89

Qy   61 NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   90 NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV 149

Qy  121 NGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYGK 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  150 NGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQFIALTDKKYGK 209

Qy  181 KDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVTLGGK 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  210 KDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVTLGGK 269

Qy  241 VAPNKDLTVKVKNQSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLN 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  270 VAPNKDLTVKVKNQSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLN 329

Qy  301 LANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLTVSNTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  330 LANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLTVSNTG 389

Qy  361 IITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLNKLVA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  390 IITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLNKLVA 449

Qy  421 TIAGEDKVVDPGSISIKSSNHGIISVNNYITAEAAGEATLTIKVGDVTKDVKFKVTTDS 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  450 TIAGEDKVVDPGSISIKSSNHGIISVNNYITAEAAGEATLTIKVGDVTKDVKFKVTTDS 509

Qy  481 RKLVSVKANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVV 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  510 RKLVSVKANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVV 569

Qy  541 TTDSGSIGTKTIGVTGNDVGEGTVHFQNGNGATLGSLYVNVTEGNVAPKNFELVSKVGQY 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  570 TTDSGSIGTKTIGVTGNDVGEGTVHFQNGNGATLGSLYVNVTEGNVAPKNFELVSKVGQY 629

Qy  601 GQSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  630 GQSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVT 689

Qy  661 GKTPGKVDIHLTKNGATAGKATVEIVQETIAIKSVNFKPVQTENFVEKKNIGTVLELEK 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  690 GKTPGKVDIHLTKNGATAGKATVEIVQETIAIKSVNFKPVQTENFVEKKNIGTVLELEK 749

Qy  721 SNLDDIVKGINLTKETQHKVRVVKSGAEQGKLYLDRNGDAVFNAGDVKLGDVTVSQTSDS 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  750 SNLDDIVKGINLTKETQHKVRVVKSGAEQGKLYLDRNGDAVFNAGDVKLGDVTVSQTSDS 809

Qy  781 ALPNFKADLYDTILTTKYTDKGTLVFKVLKDKVITSEIGSCAVHVNVLNNPNL 833
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  810 ALPNFKADLYDTILTTKYTDKGTLVFKVLKDKVITSEIGSCAVHVNVLNNPNL 862
```

```
RESULT 2
SP_BACLI
SLAP_BACLI     STANDARD;     PRT;     874 AA.
```

```
                Copyright   GenCore version 5.1.6
                Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    April 7, 2004, 17:29:55 ; Search time 18 Seconds
                                       (without alignments)
                                       2409.691 Million cell updates/sec

Title:          US-09-844-281-1
Perfect score:  4202
Sequence:       1.AGKSFPDVPAGHWAEGSINY........ITSEIGSQAVHVNVLNNPNL 833

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:        141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  SwissProt_42:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4202	100.0	862	1	SLA2_BACAA	P94217 bacillus an
2	2833.5	67.4	874	1	SLAP_BACLI	P49052 bacillus li
3	719	17.1	814	1	SLA1_BACAA	P49051 bacillus an
4	529	12.6	531	1	YQ42_BACAN	Q9rmz0 bacillus an
5	348	8.3	1176	1	SLAP_BACSH	P38537 bacillus sp
6	265.5	6.0	762	1	SLAP_ACEKI	P22258 acetogenium
7	252.5	6.0	941	1	GUN_BACS6	P19424 bacillus sp
8	230	5.5	1325	1	YDEK_ECOLI	P32051 escherichia
9	223.5	5.5	2003	1	YDBA_ECOLI	P33666 escherichia
10	219	5.2	1116	1	SLPH_BRBCH	P38538 brevibacill
11	217	5.2	1655	1	OMPB_RICCN	Q9kka3 r outer mem
12	211	5.0	1053	1	SLPM_BACBR	P05546 bacillus br
13	207.5	4.9	1645	1	OMPB_RICTY	P96989 r outer mem
14	207.5	4.9	2358	1	YEEJ_ECOLI	P76347 escherichia
15	207.5	4.9	2660	1	YEEJ_ECO57	Q8x8v7 escherichia
16	206.5	4.9	1654	1	OMPB_RICRI	Q53047 r outer mem
17	206	4.8	1698	1	41_DROME	Q9v8r9 drosophila
18	200	4.8	1643	1	OMPB_RICPR	Q53020 r outer mem
19	199.5	4.7	939	1	SLAP_CAMFE	P35827 campylobact
20	197.5	4.7	1608	1	HLYA_SERMA	P15320 serratia ma
21	195.5	4.6	1848	1	CBPA_CLOCL	P38058 clostridium
22	194.5	4.6	1656	1	OMPB_RICJA	O06653 r outer mem
23	194.5	4.6	1902	1	P2P_LACPA	Q02470 lactobacill
24	194	4.6	1300	1	120K_RICRI	P14914 rickettsia
25	193	4.6	917	1	SLAP_THETH	P35830 thermus the
26	189.5	4.5	1183	1	CNA_STRAU	Q53654 staphylococ
27	189	4.5	1953	1	BIGA_SALTY	P25927 salmonella
28	188.5	4.5	1276	1	PMP6_CHLPN	Q9z899 chlamydia p
29	187.5	4.5	1861	1	APU_THETU	P38536 t amylopull
30	185.5	4.4	1577	1	HLYA_PROMI	P16466 proteus mir
31	185	4.4	1087	1	XYNX_CLOTM	P38535 clostridium
32	183	4.4	2249	1	OMPA_RICRI	P15921 rickettsia
33	183	4.4	4705	1	FAT2_DROME	Q9vw71 drosophila
34	181.5	4.3	1228	1	SLAP_BACST	P35825 bacillus st
35	181.5	4.3	1902	1	P2P_LACLC	P15293 lactococcus
36	180	4.3	1286	1	AIDA_ECOLI	Q03155 escherichia
37	180	4.3	1569	1	YPJA_ECOLI	P52143 escherichia
38	179.5	4.3	1726	1	MSP1_PLAFC	P04934 plasmodium
39	179.5	4.3	1726	1	MSP1_PLAFP	P50495 plasmodium
40	179	4.3	1664	1	SLP1_CLOTM	Q06852 clostridium
41	178.5	4.2	2021	1	OMPA_RICCN	Q52657 rickettsia
42	176.5	4.2	447	1	ANCA_CLOTM	Q06848 clostridium
43	176.5	4.2	1902	1	P1P_LACLC	P16271 lactococcus
44	174	4.1	1025	1	SLAP_CAUCR	P35828 caulobacter
45	172	4.1	1020	1	BCA_STRAG	Q02192 streptococc

ALIGNMENTS

```
RESULT 1
SLA2_BACAA
ID   SLA2_BACAA     STANDARD;      PRT;   862 AA.
AC   P94217;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   S-layer protein EA1 precursor.
GN   EAG OR BA0887.
OS   Bacillus anthracis (strain Ames), and
OS   Bacillus anthracis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=198094, 1392;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ames;
RX   MEDLINE=22608414; PubMed=12721629;
RA   Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA   Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA   Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA   Kolonay J.F., Beanan M.J., Dodson R.J., Durkin A.S., Gwinn M.,
RA   DeBoy R.T., Madpu R., Daugherty S.C., Brinkac L.M., Haft D.H.,
RA   Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA   Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA   Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA   Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA   Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA   Fraser C.M.;
RT   "The genome sequence of Bacillus anthracis Ames and comparison to
RT   closely related bacteria.";
RL   Nature 423:81-86(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sterne / 9131;
RX   MEDLINE=97260111; PubMed=9106206;
RA   Mesnage S., Tosi-Couture E., Mock M., Gounon P., Fouet A.;
RT   "Molecular characterization of the Bacillus anthracis main S-layer
RT   component: evidence that it is the major cell-associated antigen.";
RL   Mol. Microbiol. 23:1147-1155(1997).
CC   -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
CC       of proteins which coat the surface of bacteria.
CC   -!- SUBCELLULAR LOCATION: Cell wall.
CC   -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE017027; AAP24884.1; -.
DR   EMBL; X99724; CAA68063.1; -.
DR   TIGR; BA0887; -.
DR   InterPro; IPR001119; SLH.
```

```
DR   Pfam; PF00395; SLH; 3.
KW   Signal; Repeat; Cell wall; S-layer; Complete proteome.
FT   SIGNAL        1     29       POTENTIAL.
FT   CHAIN        30    862       S-LAYER PROTEIN EA1.
FT   DOMAIN       34     76       SLH 1.
FT   DOMAIN       95    136       SLH 2.
FT   DOMAIN      157    197       SLH 3.
SQ   SEQUENCE   862 AA;  91362 MW;  CB16B202F62CCCA0 CRC64;

Query Match              100.0%;   Score 4202;  DB 1;  Length 862;
Best Local Similarity    100.0%;   Pred. No. 1.3e-175;
Matches 833;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy     1  AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    30  AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE   89

Qy    61  NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    90  NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVSAYNLKDKV  149

Qy   121  NGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQPIALTDKKYGK  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   150  NGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQPIALTDKKYGK  209

Qy   181  KDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVTLGGK  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   210  KDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVTLGGK  269

Qy   241  VAPNKQLTVKVKNQSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLN  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   270  VAPNKQLTVKVKNQSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLN  329

Qy   301  LANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLTVSNTG  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   330  LANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLTVSNTG  389

Qy   361  IITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLNKLVA  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   390  IITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLNKLVA  449

Qy   421  TIAGEDKVVDPGSISIKSSNHGIISVNNYITAEAAGEATLITIKVGDVTKDVKFKVTTDS  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   450  TIAGEDKVVDPGSISIKSSNHGIISVNNYITAEAAGEATLITIKVGDVTKDVKFKVTTDS  509

Qy   481  RKLVSVKANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVV  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   510  RKLVSVKANPDKLQVVQNKTLPVTFVTTDQYGDPFGANTAAIKEVLPKTGVVAEGGLDVV  569

Qy   541  TTDSGSIGTKTIGVTGNDVGEGTVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQY  600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   570  TTDSGSIGTKTIGVTGNDVGEGTVHFQNGNGATLGSLYVNVTEGNVAFKNFELVSKVGQY  629

Qy   601  GQSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVT  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   630  GQSPDTKLDLNVSTTVEYQLSKYTSDRVYSDPENLEGYEVESKNLAVADAKIVGNKVVVT  689

Qy   661  GKTPGKVDIHLTKNGATAGKATVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLBLEK  720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   690  GKTPGKVDIHLTKNGATAGKATVEIVQETIAIKSVNFKPVQTENFVEKKINIGTVLBLEK  749

Qy   721  SNLDDIVKGINLTKETQHKVRVVKSGAEQGKLYLDRNGDAVFNAGDVKLGDVTVSQTSDS  780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   750  SNLDDIVKGINLTKETQHKVRVVKSGAEQGKLYLDRNGDAVFNAGDVKLGDVTVSQTSDS  809

Qy   781  ALPNFKADLYDTLLTTKYTDKGTLVFKVLKDKDVITSEIGSQAVHVNLNNPNL  833
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   810  ALPNFKADLYDTLLTTKYTDKGTLVFKVLKDKDVITSEIGSQAVHVNLNNPNL  862
```

```
AC   P49052;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   S-layer protein precursor (Surface layer protein).
OS   Bacillus licheniformis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1402;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NM 105;
RX   MEDLINE=97082965; PubMed=8964497;
RA   Zhu X., McVeigh R.R., Malathi P., Ghosh B.K.;
RT   "The complete nucleotide sequence of the Bacillus licheniformis NM105
RT   s-layer-encoding gene.";
RL   Gene 173:189-194(1996).
CC   -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly
CC       of proteins which coat the surface of bacteria.
CC   -!- SUBCELLULAR LOCATION: Cell wall.
CC   -!- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
CC   --------------------------------------------------------------------------
```
```
DR   EMBL; U38842; AAC44405.1; -.
DR   PIR; JC4930; JC4930.
DR   InterPro; IPR001119; SLH.
DR   Pfam; PF00395; SLH; 3.
KW   Signal; Cell wall; S-layer; Repeat.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31    874       S-LAYER PROTEIN.
FT   DOMAIN       33     93       SLH 1.
FT   DOMAIN       94    155       SLH 2.
FT   DOMAIN      156    217       SLH 3.
SQ   SEQUENCE   874 AA;  92734 MW;  EFADCA4FF27D32AF CRC64;

Query Match              67.4%;   Score 2833.5;  DB 1;  Length 874;
Best Local Similarity    67.3%;   Pred. No. 3.6e-116;
Matches 569;  Conservative 98;  Mismatches 156;  Indels 23;  Gaps 7;

Qy     1  AGKSFPDVPAGHWAEGSINYLVDKGAITGKPDGTYGPTESIDRASAAVIFTKILNLPVDE   60
          ||||||||||||||| ||||||||||| ||||||||||||||||||||||||||||||||
Db    30  AGKSFPDVPAGHWAEDSINYLVDKGAIVGKPDGTYGPTESIDRASAAVIFTKILNLPVDE   89

Qy    61  NAQPSFKDAKNIWSSKYIAAVEKAGVVKGDGKENFYPEGKIDRASFASMLVGAYNLKEKV  120
          |||||||||||:||||||||||||||||||||:|||:||||||||||||||||||||||||
Db    90  NAQPSFKDAKNLWSSKYIAAVEKAGVVKGDGKDNFYPBGKIDRASFASMLVGAYNLKEKV  149

Qy   121  NGELVTTFEDLLDHWGEEKANILINLGISVGTGGKWEPNKSVSRAEAAQPIALTDKKYAK  180
          :| |||:|:|| | ||||||||||:||||||||||||||||||||||||||||||||||||
Db   150  DGTLVTKFDDLRGHWGEEKANILVNLGISVGTGGKWEPNKSVSRAEAAQPIALTDKKYAK  209

Qy   181  KDNAQAYVTDVKVSEPTKLTLTGTGLDKLSADDVTLEGDKAVAIEASTDGTSAVVTLGGK  240
          : ::: ::.|: ||| :|:||||:|||:|::|:||||||:||:|:|:||:|:|||||||.||
Db   210  PENSDAKVFNVAATEPTQLITLTGIGLNKLTAEDVTLEGNKAIAILEASKDGKSAVVTLSGK 269

Qy   241  VAPNKQLTVKVKNQSFVTKFVYEVKKLAVEKLTFDDDRAGQAIAFKLNDEKGNADVEYLN  300
          :||||:|.|||||.| ||:||||||||:|:||||||||||||:||:||||||||||||::|:
Db   270  IAPNKELPVKVKGNTFIVKYVYEVKKLRVEQLTFDDDRADQAVVFKLNDEKGNADIEYLD  329

Qy   301  LANHDVKFVANNLDGSPANIFEGGEATSTTGKLAVGIKQGDYKVEVQVTKRGGLTVSNTG  360
          :|.||||||||||||:|||||||||.|.|||||||||||.|.|||||||||||||:||||||
Db   330  IAGHDVKFVANNLDGTPANIFEGGTAESTTGKLAVGIABGKYKVEVQVTKRGGITVSNTG  389

Qy   361  IITVKNLDTPASAIKNVVFALDADNDGVVNYGSKLSGKDFALNSQNLVVGEKASLNKLVA  420
          ||.|||||..:|:||:|||||:|:|.||-|::|:.||||:|.|||:.::|||||.||.||||
Db   390  IIEVKNLDAEATAIKDVVFAVDTDKAG-VNTAKPLSGTDFTLNSKTIVAGEKAGIHKVVA  448
```

```
RESULT 2
P2 BACLI
SLAP_BACLI     STANDARD;     PRT;     874 AA.
```